

## **Wissenschaftliche Prüfungsarbeit**

Gemäß § 12 der Landesverordnung über die Erste Staatsprüfung für das Lehramt an Gymnasien vom 07. Mai 1982, in der derzeit gültigen Fassung

Kandidatin: Sylvie Reusch

der Johannes Gutenberg-Universität in Mainz

Fach: Französische Sprachwissenschaft

Thema: *Des idées aux mots, des mots aux idées* - L'étude de la structure et du fonctionnement d'un thésaurus français.

Erstgutachter : Prof. Dr. Bruno Staib (Johannes Gutenberg-Universität, Mainz)

Zweitgutachter : Prof. Dr. Phillipe Monneret (Université de Bourgogne, Dijon)

Abgabedatum : 06.04.2012

*À mon père*

*Toute révolution devrait commencer par une réforme de dictionnaire.*

Victor Hugo

## **Remerciements**

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma reconnaissance.

- Tout d'abord, je désire exprimer ma profonde gratitude à mes directeurs de mémoire, Philippe Monneret (Université de Bourgogne) et Bruno Staib (Johannes Gutenberg-Universität Mainz), pour leurs encadrements continus, pour les remarques constructives qu'ils m'ont fournis ainsi que pour leurs précieux conseils durant toute la période de la rédaction. Je les remercie également pour la confiance qu'ils m'ont accordée et pour la grande liberté des idées qu'ils m'ont donnée. En dehors de leurs apports scientifiques je n'oublierai pas aussi de les remercier pour leurs qualités humaines, leur hospitalité et leur soutien qui m'ont permis de mener à bien cet ouvrage.
- Je tiens également à remercier la Maison d'édition Larousse pour l'intérêt qu'il ont porté à ce travail et pour les remarques constructives qu'ils ont fait sur mon projet de mémoire.
- Je souhaite exprimer les plus grands remerciements à Daniel Péchoin pour le grand honneur de m'avoir fait partager ses expériences éditoriales avec le Thésaurus Larousse. Pendant tous les étapes de ce mémoire notre rencontre a motivé mes recherches et je rêve que notre coopération puisse avoir une digne continuation. Puis, je souhaite également remercier Milaine, son épouse, de m'avoir accueillie si chaleureusement dans leur maison.
- Finalement, je pourrais terminer cette phase de remerciements sans m'adresser à ma famille et ma belle famille pour leur soutien constant. Je leur témoigne ici de ma profonde affection et ma haute reconnaissance.

Les plus vifs remerciements s'adressent à Mathias et à mes parents, qui ont assisté quotidiennement à l'avancement de ce mémoire et qui m'ont soutenu tous les jours. Je vous remercie d'avoir accepté mes longues journées de travail, d'avoir été toujours à mon écoute et surtout de m'avoir encouragé lors des moments difficiles. Je pense également à mes beaux parents et à mon beau-frère qui m'ont soutenu maintes fois, merci beaucoup. Mes remerciements s'adressent enfin à Pascal et Susan et aux nouveaux membres de la famille, Richard et Helena, dont les naissances et les sourires sont venus apporter innocence, sens et espoir.

Enfin, je désire exprimer ma profonde gratitude à Monique Vicard et à René Heizmann pour leur disponibilité, leurs encouragements généreux et la confiance en moi qu'elles m'ont souvent transmise grâce à leur caractère chaleureux : ils m'ont été précieux.

## Table des matières

<b>Introduction .....</b>	<b>5</b>
<b>1. Le Thésaurus classique - Origine, principes et structures.....</b>	<b>8</b>
1. 1 Développement du thésaurus en France .....	8
1.1.1 À la recherche d'une forme : Entre trésor de langue et manuel des sciences .....	8
1.1.2 La découverte d'une forme : Roget's Thesaurus, modèle pour une génération .....	9
1.1.3 La remise en forme : Thésaurus Larousse et la lexicographie moderne.....	12
1.2 Conception du thésaurus français.....	14
1.2.1 Une parenté (plus ou moins) revendiquée avec Roget's Thesaurus.....	14
1.2.2 Dictionnaire idéologique : Une parenté exprimée par <i>interdépendance</i> .....	16
1.2.3 Thésaurus Larousse : Une parenté exprimée par <i>similarité organisationnelle</i> .....	23
1.3 Structure et fonctionnement du thésaurus français.....	28
1.3.1 Système de classement .....	28
1.3.2 Nature des termes.....	29
1.3.3 Nature des relations.....	30
1.3.4 Unités de classement .....	32
<b>2. Le Thésaurus Larousse - La fonction classificatoire des concepts.....</b>	<b>36</b>
2.1 Définition de l'antonymie.....	36
2.2 Analyse des quatre types d'antonymes .....	38
2.2.1 Antonymie complémentaire .....	39
2.2.2 Antonymie scalaire.....	40
2.2.3 Antonymie réciproque .....	44
2.2.4 Antonymie duale .....	46
2.3 Un classement basé sur des notions opposées.....	48
<b>3. Le Thésaurus Larousse - L'aspect relationnel des concepts .....</b>	<b>50</b>
3.1 Analyse formelle du classement par concepts.....	51
3.1.1 Définition du concept : Par l' <i>extension</i> et l' <i>intension</i> .....	51
3.1.2 Analyse formelle des concepts.....	52
3.2 Définition de la similarité sémantique et de l'analogie .....	62
3.2.1 Définition de la similarité sémantique .....	65
3.2.2 Définition de l'analogie .....	67
3.2.3 Définition de la métaphore .....	71
3.3 Analyse formelle de la similarité sémantique et de l'analogie .....	73

3.3.1 Analyse formelle de la similarité sémantique .....	73
3.3.2 Analyse formelle de l'analogie .....	75
3.4 Sémantique des cadres .....	78
3.4.1 Le cadre - structure pour la comparaison de deux concepts .....	79
3.4.2 Le cadre - représentation des connaissances du monde .....	88
3.4.3 Le Thésaurus - une conception dépendante du système de cadres .....	90
3.5 Mesure de la similarité conceptuelle dans un thésaurus .....	92
3.5.1 Le Thésaurus : Un instrument fiable pour la mesure de similarité .....	92
3.5.2 Deux approches de mesures de similarité .....	96
3.5.2.1 Approche basée sur la distance sémantique .....	97
3.5.2.2 Approche basée sur la proximité sémantique .....	103
3.5.3 Considérations sur la mesure de similarité au sein d'un thésaurus .....	115
3.5.4 La synonymie : Un cas spécifique de similarité .....	119
<b>4. Le Thésaurus numérique - Enjeux, potentiels et perspectives .....</b>	<b>123</b>
4.1 WordNet et FrameNet - Force et faiblesse des ressources numériques .....	123
4.1.1 WordNet : Un système de classement à base de <i>synsets</i> .....	123
4.1.2 FrameNet : Un système de classement à base de <i>frames</i> .....	126
4.1.3 Solution du <i>tennis problem</i> : Indicateur du succès d'un thésaurus à base de <i>frames</i> ..	128
4.2 Le Thésaurus futur - Perspectives pour une gestion des connaissances .....	130
4.2.1 Une ressource lexicale électronique constituée en <i>frames</i> .....	131
4.2.2 Un auxiliaire pour le raisonnement par analogie .....	133
4.2.3 Un appui dans la définition des concepts par <i>genus</i> et <i>differentiae</i> .....	135
4.2.4 Une structure d' <i>interlingua</i> .....	135
<b>Conclusion .....</b>	<b>138</b>
<b>Table des figures .....</b>	<b>142</b>
<b>Table des tableaux .....</b>	<b>143</b>
<b>Bibliographie .....</b>	<b>144</b>
1. Sources .....	144
2. Dictionnaires .....	144
3. Ouvrages .....	145
4. Mémoires de diplôme .....	148
5. Thèses de doctorat .....	148
6. Publications .....	149
7. Ressources numériques et manuels de thésauri .....	156

## Introduction

Le *Thésaurus* est un type de dictionnaire peu connu des français – contrairement au pays anglophones où *Roget's Thesaurus of English Words and Phrases* (1852) constitue depuis plus d'un siècle un ouvrage de consultation incontournable, renouvelé chaque année.<sup>1</sup> En France, le *Dictionnaire idéologique* (1859) de Théodore Robertson, qui constitue le pendant littéral, est aujourd'hui tombé dans l'oubli. Le *Dictionnaire idéologique* constitue la première adaptation du *classement par centres d'intérêt* à la langue française. Cependant, l'organisation logique des mots et la catégorisation rigide des concepts de l'ouvrage ne servent pas d'exemple pour la lexicographie du temps. Le *Dictionnaire idéologique* – basée sur une structure spéciale, une *hiérarchie des notions* soigneusement pensée – ne répond pas au besoin de l'imprimerie du XIXe siècle, qui songe à conquérir un marché en toute rapidité. Le *classement par concepts* va connaître son succès à travers un dictionnaire d'organisation alphabétique : Le *Dictionnaire analogique* (1862) de Prudence Boissière. L'ouvrage structure ses entrées autour des *mots-vedettes* (non autour des *concepts*), il part alors toujours du *mot* par arriver aux *idées* (non de l'*idée* pour arriver aux *mots*). La structure du 'Boissière' constitue l'archétype des dictionnaires analogiques modernes et forme leur caractère spécifique. La norme d'écriture créée cache pour longtemps le potentiel du *classement conceptuel* pour la représentation de l'analogie lexicale, qu'est la *transparence des liaisons entre idées*. Dans les années 1990, l'Éditeur Larousse décide le retour au classement conceptuel pour des raisons commerciales. L'observation du succès persistant de *Roget's Thesaurus of English Words*<sup>2</sup> inspire la Maison d'édition à rédiger un thésaurus moderne : Le *Thésaurus Larousse* (1991).

Le *Thésaurus Larousse* fonctionne selon les mêmes principes de structuration que son antécédent, le *Dictionnaire idéologique*. L'ouvrage regroupe les termes selon leur rapport de sens – indiqué par un *concept*. Le *concept* permet le repérage facile d'un *centre d'intérêt* autour duquel se forme un champ conceptuel vaste, des mots et des locutions attachés. L'ouvrage organise les concepts en fonction de leur signification (du concept plus générique au concept plus spécifique) – dans une *hiérarchie des notions*. La *hiérarchie des*

---

<sup>1</sup> Le type d'ouvrage s'intitule soit *Thésaurus* (du mot grec θησαυρός, désignant l'ouvrage de 'Trésor de langue'), soit *Dictionnaire idéologique* (du mot grec ἰδεολογία, désignant l'ouvrage de 'système classificatoire d'idées').

<sup>2</sup> Il s'agit des versions modernes, retravaillés qui ont, après plusieurs années de remaniement, peu en commun avec la structure et le lexique employé dans l'original de 1852.

*notions* permet une identification claire de la *nature du concept* et de la *nature des relations sémantiques* y attachées. Les principes structuraux définissent la nature d'un thésaurus et fondent le potentiel du classement sémantique y opéré.

La constitution d'un dictionnaire par principes de structure, autres que ceux dictés par l'ordre alphabétique, signifie pour un éditeur de tradition – tel que Larousse – une déviation par rapport à la norme. Le *Thésaurus Larousse* est longtemps considéré comme un cas d'exception – comme une sorte d'expérience commerciale. En 2009, Larousse décide une nouvelle campagne de commercialisation pour le dictionnaire. La seconde édition du *Thésaurus Larousse* <sup>3</sup> intègre les concepts existants dans l'arbitraire de l'ordre alphabétique, dissolvant les liens conceptuels et sémantiques. De ce fait, il n'existe à présent aucun thésaurus de langue française, sous forme de dictionnaire. Le *thésaurus* est devenu un champ d'investigation d'autres sciences, des sciences de l'informatique et de la documentation. Dans ces domaines scientifiques, le principe *thesaurus* constitue le fondement pour la création des grandes bases de données (comme *WordNet* et *FrameNet*) et pour l'établissement des applications informatiques du '*web sémantique*', qui visent à rendre accessible le contenu lexical des sites Web (pour des moteurs de recherche comme *Google*). Le succès du principe *thesaurus* dans divers domaines scientifiques conduit à se demander comment un dictionnaire de ce type doit être conçu pour répondre aux exigences actuelles d'innovation. La problématique à laquelle ce mémoire de Master II essaye de trouver une réponse est la suivante :

***Quel sont les fondements sur lesquels on peut élaborer un futur thésaurus de la langue française ?***

La question requiert la prise en compte des approches existantes - dans le domaine des dictionnaires et des ressources lexicales électroniques - en vue de la création d'un nouveau type de thésaurus. Pour répondre à la problématique, la mémoire adopte la démarche suivante :

Dans un premier temps, on étudie la conception du thésaurus classique par rapport à son modèle lexicographique, le *Thesaurus of English Words and Phrases*. Le Dictionnaire idéologique et le *Thésaurus Larousse* adoptent - de manière différente - le classement proposé par Peter M. Roget. Une analyse comparative des deux ouvrages (de leur structure

---

<sup>3</sup> La différence de conception par rapport au *Thésaurus Larousse* de 1991 et souligné par un changement de nom. La seconde édition apparaît en 2009 sous le titre *Dictionnaire des analogies*.

et de leur fonctionnement) permet de préciser une démarche pour la conceptualisation d'un thésaurus futur, sur base d'un dictionnaire préexistant.<sup>4</sup>

Puis, on étudie deux aspects (ou fonctions) qui sont toujours présents lors de la construction d'un thésaurus, à savoir la *fonction classificatoire des concepts* et l'*aspect relationnel des termes*. La conception du Thésaurus Larousse sert d'exemple. Dans un deuxième temps, on se concentre sur la *fonction classificatoire des concepts*. Un thésaurus est constitué d'une série des mots-clefs (descripteurs) auxquels on attribue, de manière univoque et réciproque, une notion conceptuelle (signification normalisée). Un thésaurus se construit à l'aide d'un *Plan de classement* au sein duquel deux descripteurs sont placés (si possible) en paires de notions opposées. Ceci permet de rapprocher visiblement sens et sens opposé d'un concept dans l'ossature du dictionnaire. Dans un troisième temps, on étudie l'*aspect relationnel des termes*. Un thésaurus permet à l'utilisateur de transformer ses idées en mots et de métamorphoser ses mots en idées. Ces allers-retours peuvent se réaliser uniquement dans un *réseau sémantique* qui englobe les relations de similarité sémantique, d'analogie, de et de synonymie. Le réseau sémantique est la structure sous-jacente dans chaque thésaurus qui permet de rapprocher les termes dans une perspective sémantique. La connaissance des deux aspects (ou fonctions) d'un thésaurus - qui sont strictement enchaînés - seront d'importance en vue de la création d'un système de classement au sein d'un futur thésaurus.

Dans un quatrième temps, on étudie la conception de deux ressources lexicales électroniques, WordNet et FrameNet. La dynamique du réseau lexical constitue l'innovation de son classement sémantique. De multiples applications soutiennent *activement* la recherche et l'interconnexion des informations lexicales. Dans la perspective de la création d'un futur thésaurus, l'appui sur une structure informatique établie et communément acceptée - comme celle de WordNet - semble attractif. La tendance à vouloir adopter un modèle lexicographique prospère se réaffirme. Une analyse contrastive avec le Thésaurus Larousse sert - dans cette perspective - à clarifier les avantages et les limites d'une telle organisation lexicale électronique. Le mémoire conclut sur un aperçu de ce qui pourrait constituer le futur thésaurus de la langue française. Cet aperçu est le fondement pour la réalisation d'un prototype dans un travail ultérieur.

---

<sup>4</sup> Les analyses dans le présent Mémoire s'appuieront sur les publications suivants : Dictionnaire idéologique (1859), Thesaurus of English Words (1856 ; 1911 ; 1987 ; 2002) et Thésaurus Larousse (1992).

# 1. Le Thésaurus classique - Origine, principes et structures

## 1. 1 Développement du thésaurus en France

### 1.1.1 À la recherche d'une forme : Entre trésor de langue et manuel des sciences

À la Renaissance, le thésaurus (du mot grec *θησαυρός* : trésor) est un dictionnaire de langue ancienne - de préférence latine - visant l'exhaustivité du lexique de la langue et l'explication de chaque mot par son contexte sémantique.<sup>5</sup> Le classement méthodique adopté a une longue tradition qui remonte jusqu'aux premiers *vocabulaires* médiévaux en Angleterre qui s'organisent autour de centres d'intérêt.<sup>6</sup> En France, la diffusion des thésauri est étroitement liée à l'Imprimerie. Les frères Henri et Robert Estienne diffusent les premiers exemplaires : Le *Thesaurus lingua latinae* (1531), le *Dictionarium Latino-Gallicum* (1538) et le *Dictionnaire françois-latin* (1540).<sup>7</sup> Le classement par centres d'intérêt trouve sa fin avec le développement de la lexicographie française. Le *Thresor de la langue françoise* (1606) de Jean Nicot définit pour les années à venir le caractère du dictionnaire monolingue.<sup>8</sup> Durant le XVII<sup>e</sup> et XVIII<sup>e</sup> siècle, le Dictionnaire par ordre alphabétique devient le prototype de l'Imprimerie royale et de l'Académie française - les grandes institutions linguistiques du temps. Au XIX<sup>e</sup> siècle, l'essor des Sciences naturelles met en cause la classification alphabétique des connaissances. La production de nombreuses nomenclatures telles que celles de Lavoisier pour la Chimie et de Linné et Jussieu pour la Botanique, avaient suscités chez les lexicographes l'ancien désir d'entreprendre une classification exhaustive des mots, des symboles qui expriment les connaissances du monde et les expériences de l'homme.<sup>9</sup> En 1853, A. de Ponton d'Amécourt publie le *Panorama des mots, Nouveau dictionnaire des Synonymes*. Il présente la première tentative d'un *classement idéologique* en France.<sup>10</sup> Ponton d'Amécourt est un grand admirateur des nomenclatures scientifiques. Dans sa Préface il se

---

<sup>5</sup> Cf. Marelllo, Encyclopédie internationale de lexicographie, Tome I, article : The Thesaurus, p. 1083.

<sup>6</sup> Cf. Quemada (1967), p.361.

<sup>7</sup> Cf. Nativel (1997), p.341 ; cf. Matoré (1968), pp. 59/60.

<sup>8</sup> Cf. Bloch (1904), pp. 1-13 ; cf. Matoré (1968), pp. 59/60.

<sup>9</sup> Cf. Quemada (1967), pp. 377/378 ; cf. Hüllen (2009), p. 89 ; Cf. Roget's Thesaurus (2002), Roget's Preface to the first edition, pp. xxiii-xix ; cf. Dictionnaire idéologique (1859), Introduction, pp. iii-iv ; cf. Dictionnaire logique de la langue française (1882), Tome III. Représentation logique, p.2 (en ressource numérique).

<sup>10</sup> Cf. Péchoin (2001), p. 97 ; cf. Quemada (1967), p. 386 ; cf. Rey (1973), p. 168.

L'*Idéologie* est vue comme un aboutissement de la philosophie classique de la représentation. Dans le cadre de la sémiotique généralisant le *classement idéologique* devient une approche lexicographique, selon laquelle les signes du langage, quelle que soit leur importance, sont subordonnés à l'ordre préalable de la conscience.



réfère explicitement à eux, déplorant qu'aucun linguiste n'ait tenté de créer une pareille classification pour les *Sciences de la langue et de la pensée*<sup>11</sup> :

« La zoologie la botanique, la minéralogie, les sciences physiques comme les sciences mathématiques, tout cela a été exploré par de laborieux inquisiteurs qui ont légué d'admirables classifications ; mais ce qu'on a fait pour la nature inerte (...) nul n'avait songé à le faire pour l'intelligence qui vit et crée, pour la pensée de reine du monde (...), le plus noble apanage de l'homme. On étudiait théorétiquement les phénomènes psychologiques, mais pour en venir à des résultats d'une utilité pratique, pour chercher à classer les idées et les mots qui les expriment dans toutes les langues, dresser une carte géographique de l'esprit humain, étiqueter, en quelque sorte, toutes les pensées, déterminer les familles, comparer les idiomes, (...) élargir la sphère de l'intelligence (...) nul n'y avait songé. »<sup>12</sup>

Le *Panorama des mots* constitue une première approche de la classification par centres d'intérêt, profondément inspirée par les sciences naturelles. L'ouvrage se compose de deux parties. Un premier livre sert au classement des « Objets de mathématiques, de la physique, de l'histoire naturelle et de la physiologie », un second livre au classement de « Tout ce qui concerne les personnes ». <sup>13</sup> Le *Panorama des mots* se donne pour une classification psychologique du langage. <sup>14</sup> « [Il présente] un vaste tableau synoptique ou les éléments de [la] pensée [ont] tous été rangés logiquement avec les formes nombreuses que le langage leur a données chez tous les peuples et dans tous les siècles ». <sup>15</sup> Le projet n'est réalisé que partiellement et meurt avec son auteur. Le *Panorama des mots* reste donc un ouvrage inachevé. <sup>16</sup> En même temps, paraît en Grande Bretagne la première édition du *Thesaurus of English Words and Phrases* (1852). <sup>17</sup>

### 1.1.2 La découverte d'une forme : Roget's Thesaurus, modèle pour une génération

Le *Thesaurus of English Words and Phrases* va devenir l'archétype du thésaurus moderne en Europe. <sup>18</sup> L'auteur, Peter Mark Roget (1779 - 1869) est un professeur de médecine à Université de Manchester et Membre de la Société Royale de Londres. <sup>19</sup> Pour son travail Peter Mark Roget élabore des listes de mots et de locutions, qu'il classe dans divers ordres d'idées dans un agenda, pour pouvoir s'exprimer de manière soignée et diversifié sur un

---

<sup>11</sup> Cf. Quemada (1967), p. 386.

<sup>12</sup> Cf. Quemada (1967), p. 386 ; cf. *Panorama des mots* (1853), Préface, pp. ii-iii.

<sup>13</sup> Cf. Quemada (1967), pp. 387-389.

<sup>14</sup> Cf. Péchoin (2001), p. 97.

<sup>15</sup> Cité d'après Péchoin (2001), p. 97.

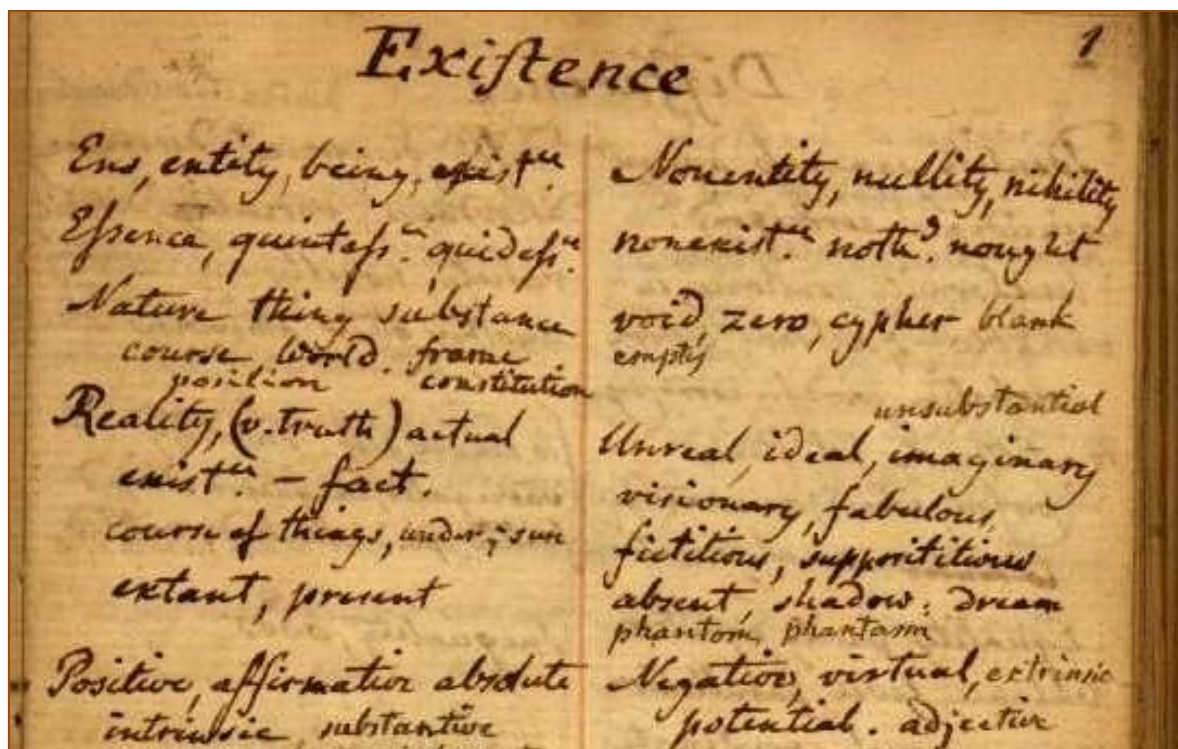
<sup>16</sup> Cf. Péchoin (2001), p. 89.

<sup>17</sup> Cf. Marengo, Encyclopédie internationale de lexicographie, Tome I, article : The Thesaurus, p. 1083.

<sup>18</sup> Cf. Péchoin (2001), p. 89.

<sup>19</sup> Cf. Péchoin (2001), p. 89 ; cf. *Roget's Thesaurus* (2002), Dr Peter Mark Roget and his Thesaurus, pp.xi/xii.

nouveau sujet.<sup>20</sup> La figure 1 montre une page de carnet, portant sur la notion de l'Existence.



**Figure 1: Manuscrit de Peter Mark Roget**<sup>21</sup>

En 1848, Roget se retire de sa carrière médicale et s'intéresse à la réalisation de l'ouvrage dont il avait depuis si longtemps commencé à réunir les matériaux.<sup>22</sup> Le thésaurus présente pour Roget le système cohérent dans lequel il range sa collection de mots et de locutions pour la mettre à disposition aux savants de son temps.<sup>23</sup> Dans l'introduction au dictionnaire Roget précise l'objectif du projet:

« The present work is intended to supply, with respect to the English language, a desideratum hitherto unsupplied in any language; namely, a collection of words it contains and of the idiomatic combinations particular to it, arranged, not in alphabetical order as they are in a Dictionary, but according to the *ideas* which they express. (...) The idea being given, [we are] to find the word, or words, by which that idea may be most fitly and aptly expressed. For this purpose, the words and phrases (...) are here classed (...) strictly according to their signification. »<sup>24</sup>

<sup>20</sup> Cf. Marello, Encyclopédie internationale de lexicographie, Tome I, article : The Thesaurus, p. 1083. cf. Roget's Thesaurus (2002), Preface to the first edition, 1852, p. xviii ; cf. Roget's Thesaurus (2002), The Éditeurs Préface de George Davidson, p. vii.

<sup>21</sup> Figure 1 : Adopté du site internet <http://www.rain.org/~karpeles/rogetdis.html>

<sup>22</sup> Cf. Old (2003), p. 9.

<sup>23</sup> Cf. Roget's Thesaurus (2002), Dr Peter Mark Roget and his Thesaurus, p. xiv.

<sup>24</sup> Cité d'après Roget's Thesaurus (2002), Introduction to the first edition. 1852, p.xix.

Le thésaurus devient aux mains de Roget un outil qui aide à trouver à partir d'une idée donnée le mot qui l'exprime le plus convenablement.<sup>25</sup> En 1859, Le *Dictionnaire idéologique* de Théodore Robertson est le « premier essai d'application cette formule lexicographique » en français.<sup>26</sup> Théodore Robertson est un professeur d'anglais qui est vivement intéressé dans la rédaction de dictionnaires et de manuels d'enseignement. Le *Nouveau cours pratique, analytique, théorétique et synthétique de langue anglaise* (1838) et *Dictionnaire des racines anglaises, et traité de la formation des mots* (1839) comptent parmi ses premières publications.<sup>27</sup> En 1852, lors de la publication du *Thesaurus of English Words and Phrases*, Robertson est occupé à chercher des matériaux pour un nouveau dictionnaire de langue. Il confesse dans la Préface au Dictionnaire idéologique :

« J'avais employé des années à réunir et à classer des matériaux. En examinant l'ouvrage de M. Roget, je reconnus que le recueil, qui devait servir de base à mon travail, était déjà tout fait : qu'il était plus complet que celui qui m'avait coûté tant de soins et qu'il était bien supérieur sous rapport de classification.(...) Frappé de l'utilité du Thesaurus, j'avais résolu de suspendre tout autre occupation pour appliquer à la langue française le plan suivi dans cet excellent ouvrage. »<sup>28</sup>

Le *Plan of classification* adopté par Peter Mark Roget est jugé d'un degré de généralité suffisant pour être appliqué à toutes les langues européennes afin d'établir un lexique polyglotte.<sup>29</sup> Pour Théodore Robertson, le *Dictionnaire idéologique* en est la première démonstration. Joint au travail effectif de la réalisation d'un thésaurus français, Robertson voit dans l'analyse du *Thesaurus of English Words and Phrases* un chemin vers la détermination des principes d'une langue universelle<sup>30</sup>. Ainsi, il écrit dans sa préface :

« [...] ce n'est qu'à l'aide des semblables analyses que l'on pourra déterminer les principes d'une langue rigoureusement philosophique. Le résultat probable de la création d'une telle langue serait son adoption définitive pour toutes les nations civilisées ; réalisant ainsi ce magnifique rêve des philanthropes, l'établissement d'une langue universelle. »<sup>31</sup>

L'existence d'une telle 'langue universelle' permettrait une véritable connaissance des éléments qui entrent dans la formation des idées, les actes du raisonnement et la communication de la pensée.<sup>32</sup> Au XIX<sup>e</sup> siècle, ces ambitions philosophiques causent le

<sup>25</sup> Cf. Péchoin (2001), pp.89/90.

<sup>26</sup> Cité d'après Quemada (1967), p. 380.

<sup>27</sup> Cf. Hüllen (2009), p. 77.

<sup>28</sup> Cité d'après Dictionnaire idéologique (1859), Préface, p. i.

<sup>29</sup> Cf. Quemada (1967), p. 381.

<sup>30</sup> Cf. Quemada (1967), p. 381.

<sup>31</sup> Cité d'après Dictionnaire idéologique (1859), Introduction, p. xiii.

<sup>32</sup> Cf. Quemada (1967), p. 380.

grand succès du *Thesaurus of English Words and Phrases*. Bien que l'ouvrage soit lui-même le résultat d'une longue et complexe tradition, la forme particulière dans laquelle il apparaît en 1852 est tellement convaincante - surtout dans un contexte d'ethnocentrisme - qu'après une centaine d'années presque toutes les langues européennes ont leur propre thésaurus.<sup>33</sup> Le *Plan of classification*, créé par Roget, va servir de guide à de nombreux ouvrages dont pour la langue allemande, les dictionnaires de August Schlessing (1881) et de Daniel Sanders (1873).<sup>34</sup> Au XXe siècle, les lexicographes se détachent du projet philosophique et commencent à utiliser leurs propres systèmes de catégories.<sup>35</sup> Les dictionnaires de Franz Dornseiff (1933) et de Rudolf Hallig & Walther von Wartburg (1952) sont les témoins d'un système de classification d'idées développé pour la langue allemande.<sup>36</sup>

### 1.1.3 La remise en forme : Thésaurus Larousse et la lexicographie moderne

Au XXe siècle, au temps de la lexicographie commerciale, les considérations de nature philosophique demeurent à l'arrière-plan, tandis que s'affirment des préoccupations plus concrètes en vue de l'utilisation du thésaurus.<sup>37</sup> L'éditeur Larousse songe avec la publication du *Thésaurus Larousse* (1999) à proposer de manière commode aux professionnels de la communication des expressions dont ils ont besoin et en même temps leur donner, par le moyen de termes analogues, une gamme d'association d'idées très large, susceptible d'élargir le champ de leur imagination créatrice.<sup>38</sup> Le *Thésaurus Larousse* se présente comme une adaptation moderne du modèle lexicographique de Peter Mark Roget. Le directeur de l'ouvrage, Daniel Péchoin, décrit son travail avec le *Thesaurus of English Words and Phrases* de la façon suivante :

« C'est la vision globale qui nous a intéressé dans le dictionnaire idéologique. Car en proposant à Larousse de réaliser un thésaurus à la manière de Roget, nous voulions simplement rendre plus facile à consulter et plus stimulant pour l'esprit du lecteur un ouvrage appelé à jouer le rôle à la fois d'un dictionnaire de synonymes, d'un dictionnaire des antonymes et d'un dictionnaire analogique : nous ne rêverions ni de mettre au point un « classement ontologique » absolu [...], ni de « déterminer les principes d'une langue rigoureusement philosophique » susceptible d'être adoptée « par toutes les nations civilisées » comme Robertson, ni de « classer les

<sup>33</sup> Cf. Hüllen (2009), p. 89.

<sup>34</sup> Cf. Hüllen (2009), p. 89.

<sup>35</sup> Cf. Hüllen (2009), p. 60.

<sup>36</sup> Cf. Hüllen (2009), p. 89 ; cf. De Tollenaere (1960), p. 20.

<sup>37</sup> Cf. Quemada (1967), pp. 380/381.

<sup>38</sup> Cf. Quemada (1967), pp. 380/381 ; cf. *Thésaurus Larousse* (1992), Préface, p. v.

idées et les mots qui les expriment dans toutes les langues » comme Ponton d'Amécourt. »<sup>39</sup>

Les créateurs du *Thésaurus Larousse* estiment les exigences extralinguistiques déterminantes pour la conceptualisation d'un système de catégories.<sup>40</sup> Ces exigences<sup>41</sup> déterminent comment chaque langue opère un découpage différent de la réalité. Cette nouvelle perspective sur les langues se développe dans la perspective de la linguistique structurale, puis de la linguistique fonctionnelle, entre les années 1920 et 1960, comme l'explique Daniel Péchoin :

« Depuis le temps de ces aïeux [...] l'eau de la linguistique a passé sous le pont des universaux. Saussure, puis les anthropologues linguistes américains des premières décennies de ce siècle nous ont fait nous déprendre de certaines illusions et en particulier de « l'idée simpliste que le monde tout entier s'ordonne antérieurement à la vision qu'en ont les hommes, en catégories d'objets parfaitement distinctes, chacune recevant nécessairement une désignation dans chaque langue »<sup>42</sup>, selon la formule d'André Martinet : n'en déplaise à Platon et aux réalistes universaux qui voulaient, tel le boucher « découper [la réalité] aux jointures, comme il lui convient par nature »<sup>43</sup>, nous savons aujourd'hui qu'on ne découpe pas la réalité de la même façon en anglais et en français, comme le montre l'irréductible non-coïncidence, à la fois pragmatique et conceptuelle, du *T-bone steak* et de la côte de bœuf. »<sup>44</sup>

Le *Thésaurus Larousse* regroupe les concepts selon une vision du monde dictée par le français.<sup>45</sup> Le *système des catégories* de Peter Mark Roget sert aux lexicographes de grille pour la structuration, les *notions* servent d'*étiquettes* pour la conceptualisation de l'ouvrage.

Le précis historique montre que le *Thesaurus of English Words and Phrases* est considéré pendant plus d'un siècle comme le véritable promoteur d'une nouvelle formule lexicographique. Une formule qui permet aux lexicographes « une perception claire des rapports qui existent entre [...] symboles et [...] idées correspondantes »<sup>46</sup> lors de la conception d'un thésaurus français.

---

<sup>39</sup> Cité d'après Péchoin (2001), p. 117.

<sup>40</sup> Cf. Quemada (1967), p. 382.

<sup>41</sup> exigences extra-linguistique: influences culturelles, géographiques, etc.

<sup>42</sup> Péchoin cite ici Martinet (1967), p.11.

<sup>43</sup> Péchoin cite ici Martinet (1967), p.11.

<sup>44</sup> Cité d'après Péchoin (2001), pp. 117/118.

<sup>45</sup> Cf. *Thésaurus Larousse* (1992), Préface, pp. vi/vii.

<sup>46</sup> Cité d'après Dictionnaire idéologique (1859), Introduction, p. xiii.

## 1.2 Conception du thésaurus français

### 1.2.1 Une parenté (plus ou moins) revendiquée avec Roget's Thesaurus

Le *Thesaurus of English Words and Phrases* est une grande source d'inspiration pour les lexicographes en France. Cependant, il existe des différences de conception entre le *Dictionnaire idéologique* et le *Thésaurus Larousse* et leur ancêtre commun. Au niveau des informations conceptuelles et lexicales encodées, les ouvrages présentent différents degrés de parenté au modèle lexicographique anglais. Deux formes de parenté peuvent être distinguées : l'interdépendance (a) et la similarité organisationnelle (b)<sup>47</sup> :

#### a) La similarité organisationnelle

La similarité organisationnelle concerne deux ouvrages qui partagent le même principe d'organisation. Un ouvrage est conçu *d'après* le modèle de l'autre et reconnaît ainsi les mêmes entités structurelles (les catégories, les divisions, etc.).<sup>48</sup>

#### b) L'interdépendance

L'interdépendance concerne deux ouvrages qui partagent le même contenu conceptuel et lexical. Un ouvrage est conçu *sur* le modèle de l'autre. L'interdépendance inclut ainsi la notion d'*alignement* qui indique l'identité de deux ouvrages à plusieurs niveaux hiérarchiques.<sup>49</sup>

Par rapport au *Thesaurus of English Words and Phrases*, le *Dictionnaire idéologique* se trouve dans une relation d'interdépendance et le *Thésaurus Larousse* dans une relation de similarité organisationnelle<sup>50</sup> Les deux *degrés de parenté* caractérisent deux types de thésaurus, à l'égard de la manière dont un ouvrage est conceptualisé par rapport au thésaurus anglais. Deux approches peuvent être distinguées : L'approche par expansion (a) et l'approche par dérivation (b).<sup>51</sup>

---

<sup>47</sup> Cf. Lönneker-Rodman (2007), p.2.

<sup>48</sup> Cf. Lönneker-Rodman (2007), p.2 ; cf. Roget's Thesaurus (2002), A selection of pages to the first edition : Plan of Classification, p. xxxiv ; cf. Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.

<sup>49</sup> Cf. Lönneker-Rodman (2007), p.3. cf. Roget's Thesaurus (2002), A selection of pages to the first edition : Plan of Classification, p. xxxiv ; cf. Dictionnaire idéologique (1859), Tableaux synoptiques des catégories, pp. xiv-xxvii.

<sup>50</sup> Cf. Hüllen (2009), pp. 84-89 ; cf. Péchoin (2001), pp. 101/102 et 117/118.

<sup>51</sup> Cf. Lönneker-Rodman (2007), p.3.

a) L'approche par expansion

Dans une approche par expansion on transfère un modèle de thésaurus, considéré comme avantageux, à une autre langue. Dans notre cas de figure, on établit un thésaurus français sur le modèle anglais préexistant. La notion d'*expansion* implique que la structure globale (de catégories et d'articles) du thésaurus source reste inchangée lors du transfert vers la langue cible, bien qu'on ajoute quelques unités lexicales inhérentes au français, au sein d'entrées lexicales. L'approche expansive tend à reproduire deux thésauri structurellement très similaires. Ceci peut entraîner le risque qu'on néglige les spécificités de lexicalisation d'une langue, dans la perspective d'une conceptualisation cohérente de l'ouvrage cible.<sup>52</sup>

b) L'approche par dérivation

Dans une approche par dérivation on se rattache à un modèle de thésaurus préexistant dans une autre langue, sans s'approprier sa conceptualisation inhérente. Dans notre cas de figure, un thésaurus français va présenter une structure divergente du *Thesaurus of English Words and Phrases*, car les lexicographes se réservent la liberté de prendre des décisions en faveur d'un classement lexical dicté par la langue cible. La notion de *dérivation* implique ainsi une évaluation critique du modèle lexicographique lors de l'alignement des ressources. L'approche dérivationnelle tend à reproduire deux thésauri structurellement divergents, ceci peut entraîner des risques pour une comparaison subséquente des deux ouvrages, car un alignement en vue d'une convergence s'avère complexe.<sup>53</sup>

Les différences de conception entre le *Dictionnaire idéologique* et le *Thésaurus Larousse* s'expliquent par leur démarche singulière lors de l'adaptation du modèle lexicographique anglais. Lors de l'établissement du *Dictionnaire idéologique* on opte en faveur d'une approche par expansion, lors de l'établissement du *Thésaurus Larousse* en faveur d'une approche par dérivation. Ces démarches lexicographiques s'expliquent par un contexte historique précis qui détermine l'*usage* des dictionnaires. L'analyse du système classificatoire des deux ouvrages s'accompagne donc des réflexions sur les exigences historiques qui influencent la conception du thésaurus français.

---

<sup>52</sup> Cf. Lönneker-Rodman (2007), pp.4/5 ; cf. Hüllen (2009), pp. 84-89 ; cf. Dictionnaire idéologique (1859), Introduction, pp. xii/xiii.

<sup>53</sup> Cf. Lönneker-Rodman (2007), p.4 ; cf. Péchoin (2001), pp. 117/118 ; cf. Thésaurus Larousse (1992), Préface, pp. vi/vii ; cf. Péchoin (2001), pp. 117/118.

### 1.2.2 Dictionnaire idéologique : Une parenté exprimée par *interdépendance*

Théodore Robertson considère que le système de classification du *Thesaurus of English Words and Phrases* peut être utilisé pour mettre en relation différentes langues dans un objectif de comparaison ou d'appréhension :

« Le *plan de classification* adopté dans cet ouvrage étant applicable à toutes les langues, vivantes ou mortes, on pourrait composer de la même manière un dictionnaire allemand, italien, espagnol, latin ou grec, possédant tous les avantages qu'offre le modèle anglais. Plus utile encore serait la combinaison de ces compilations méthodiques appliquées à deux langues, le français et l'anglais, par exemple [...]. Pour les étudiants de chacune des deux nations, ce serait un excellent moyen de se perfectionner dans la langue de l'autre. [...] Pour les philologues, ce serait le moyen le plus facile et le plus efficace d'établir une comparaison exacte entre deux langues, et d'apprécier justement leurs mérites et leurs défauts respectifs. On obtiendrait une somme d'avantages plus grande encore, en construisant un *Lexique polyglotte* d'après ce système. »<sup>54</sup>

Dans cette perspective, le *Dictionnaire idéologique* (1959) constitue le prototype pour la transposition du système Roget vers d'autres langues européennes en vue de la création d'un lexique polyglotte.<sup>55</sup> L'ouvrage se construit dans une *approche par expansion* par rapport au *Thesaurus of English Words and Phrases*. Une comparaison détaillée des deux ouvrages montre que leur structure classificatoire est *quasi* identique : Le '*Plan of classification*' et le '*Plan de classification*' sont symétriques (cf. tableau 1).

---

<sup>54</sup> Cité d'après Dictionnaire idéologique (1859), Introduction, pp. xii/xiii.

<sup>55</sup> Cf. Quemada (1967), p. 381.



**Tableau 1: Comparaison de la structure classificatoire entre Roget et Robertson** <sup>56</sup>

<u>Roget (1852)</u>			<u>Robertson (1859)</u>		
<b>Plan of classification</b>			<b>Plan de classification</b>		
<b>I. Abstract Relations</b>	1. Existence	2. Relation	<b>I. Rapports abstraits</b>	1. Existence	2. Relation
	3. Quantity	4. Order		3. Quantité	4. Ordre
	5. Number	6. Time		5. Nombre	6. Temps
	7. Change	8. Causation		7. Changement	8. Causalité
<b>II. Space</b>	1. Generally	3. Form	<b>II. Espace</b>	1. En général	2. Forme
	2. Dimensions	4. Motion		2. Dimensions	4. Mouvement
<b>III. Matter</b>	1. Generally	2. Inorganic	<b>III. Matière</b>	1. En général	2. Inorganique
	3. Organic			3. Organique	
<b>IV. Intellect</b>	1. Formation of ideas	2. Communication of ideas	<b>IV. Intelligence</b>	1. Formation des idées	2. Communication des idées
<b>V. Volition</b>	1. Individual	2. Intersocial	<b>V. Volonté</b>	1. Individuelle	2. Mutuelle
<b>VI. Affections</b>	1. Generally	2. Personal	<b>VI. Affections</b>	1. En général	2. Personnelles
	3. Sympathetic	4. Moral		3. Sympathiques	4. Morales
	5. Religious			5. Religieuses	

La seule divergence entre les tableaux classificatoires est de nature *stylistique*. Alors que Peter Mark Roget utilise des *mots anglais latinisés* pour désigner les classes conceptuelles, Théodore Robertson opte pour une traduction directe (cf. tableau 2). <sup>57</sup>

**Tableau 2: Divergences entre les tableaux classificatoires de Roget et Robertson** <sup>58</sup>

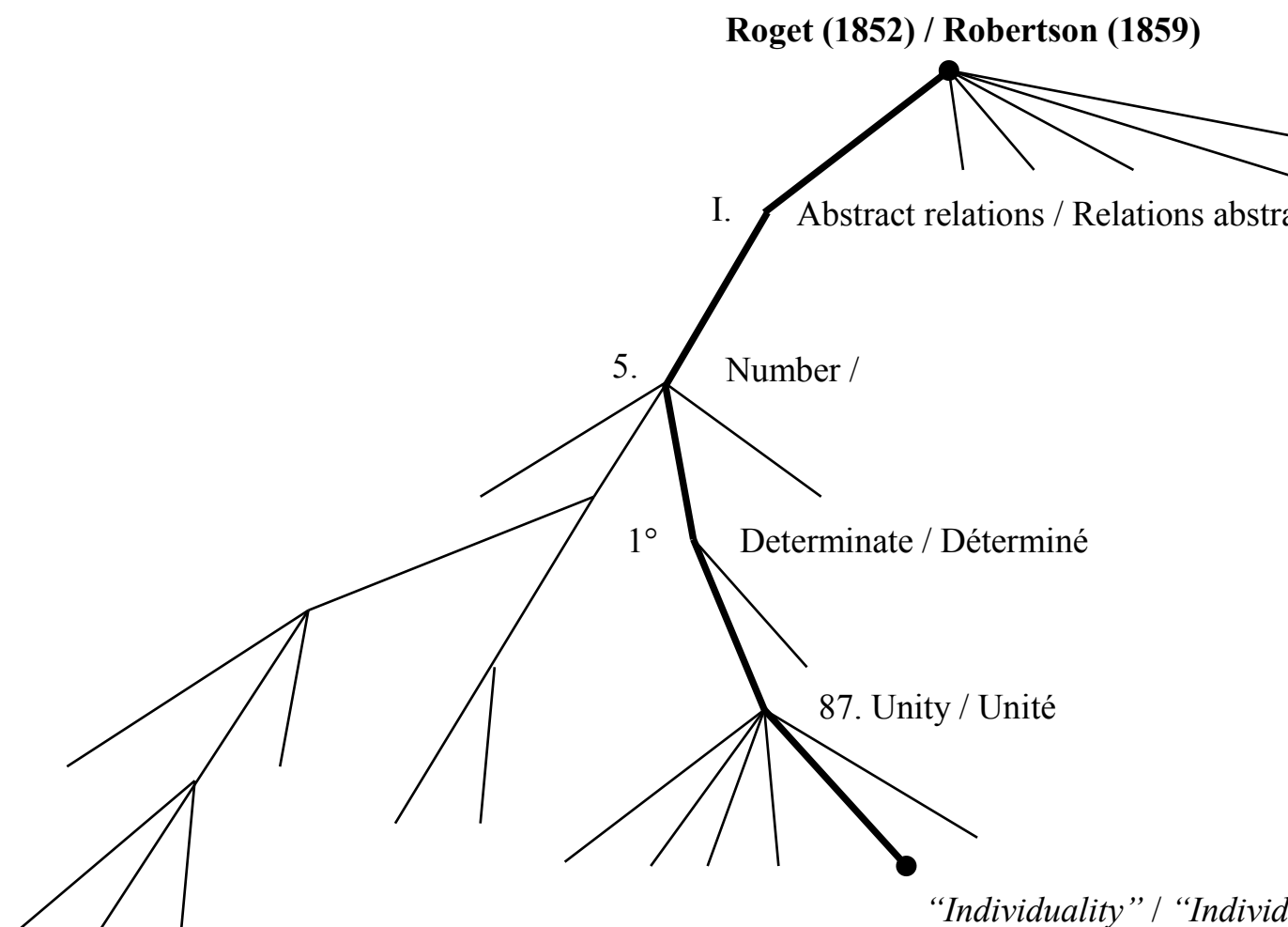
<b>Roget</b>	<b>Robertson</b>
<i>V. Volition</i> [will : volonté]	<i>V. Volonté</i> [volition: volition]
<i>2. Intersocial</i> [mutel : mutelle]	<i>2. Mutuelle</i> [intersocial : intersocial]

<sup>56</sup> Tableau 1 de S. Reusch, élaboré d'après les informations trouvés dans Roget's Thesaurus (2002), A selection of pages to the first édition : Plan of Classification, p. xxxiv et du Dictionnaire idéologique (1859), Plan de classification, p.xiv.

<sup>57</sup> Cf. Hüllen (2009), p.85.

<sup>58</sup> Tableau 2 de S. Reusch, élaboré d'après les informations trouvés Hüllen (2009), pp. 85/86.

Le *Plan of classification* sert de modèle au classement lexical opéré au sein du *Dictionnaire idéologique*. Le parallélisme structurel entraîne que la recherche d'informations s'attache au même *chemin de consultation* dans les deux ouvrages (cf. figure 2).<sup>59</sup>



**Figure 2: Le chemin de consultation identique chez Roget et Robertson**

La figure 2 montre un exemple pour la recherche de l’item ‘*individualité*’ (angl. ‘*individuality*’). L’item constitue un mot d’entrée dans l’article 87. *Unité* (*Unity*), qui figure au Plan de classification du *Thesaurus of English Words and Phrases* et du *Dictionnaire idéologique*. Dans les deux ouvrages, le classement sémantique situe cet article dans la catégorie I. *Rapports abstraits* (*Abstract relations*), la division 5. *Nombre* (*Number*) et la sous-division 1° *déterminé* (*determinate*). Le chemin pris lors de la

<sup>59</sup> Figure 2 de S. Reusch, élaboré d’après les informations trouvés dans le Dictionnaire idéologique (1859), Tableaux synoptiques des catégories, pp. xvi-xxvii, dans Roget’s Thesaurus (2002), A selection of pages to the first édition : Plan of Classification, p. xxxiv et dans Hüllen (2009), p. 85.

recherche de l’item est *identique* dans les deux ouvrages. L’exemple est emblématique pour la totalité des 1000 articles.

Dans l’ambition épistémologique d’une *synchronisation des ouvrages*, Robertson ne transpose pas uniquement le classement sémantique d’après le modèle du *Plan of classification*, comme le montre l’organisation des articles 87. *Unité*, 89. *Dualité*, 92. *Trinité* et 94. *Quaternity* dans les deux ouvrages (en encadré dans la figure suivante) <sup>60</sup> ...

**Tableau 3 : Interdépendance au niveau de la macrostructure** <sup>61</sup>

Thesaurus of English Words (1856)		Dictionnaire idéologique (1859)
CLASS I : ABSTRACT RELATIONS		CLASSE I : RAPPORTS ABSTRAITS
V. Number		V. Nombre
1. Abstract		1. Abstrait
84. Number		84. Nombre
85. Numeration		85. Numération
86. List		86. Liste
2. Determinate		2. Déterminé
<div style="border: 1px solid black; padding: 5px; display: inline-block;"> <b>87. Unity</b>  <b>89. Duality</b>            90. Duplication  <b>92. Trinity</b>            93. Triplication  <b>94. Quaternity</b> </div> 88. Accompani- { -ment 91. Bisection 94. Trisection	→	<div style="border: 1px solid black; padding: 5px; display: inline-block;"> <b>87. Unité</b>  <b>89. Dualité</b>            90. Duplication  <b>92. Trinité</b>            93. Triplicité  <b>94. Quaternité</b> </div> 88. Accompagne- { -ment 91. Bisection 94. Trisection
...		...
105. Infinity		105. Infinité

... mais Robertson transpose aussi la *structure* des entrées lexicales du *Thesaurus of English Words and Phrases*, comme le montre l’organisation des entrées lexicales de l’article 87. *Unity* (indiqué par un soulignage) ...

<sup>60</sup> Cf. Hüllen (2009), pp. 86-89.

<sup>61</sup> Tableau 3 de S. Reusch, élaboré d’après les informations trouvés dans Hüllen (2009), pp. 86-89, dans le Dictionnaire idéologique (1859), Tableaux synoptiques des catégories, pp. xvi-xxvii et dans Roget’s Thesaurus (2002), A selection of pages to the first édition : Plan of Classification, p. xxxiv

**Tableau 4 : Interdépendance au niveau de la microstructure** <sup>62</sup>

**Thesaurus of English Words (1856)**

**87 UNITY.** *N.* unity; oneness.; individuality; singleness, solitariness, solitude 893; isolation, abstraction.

One, unit, ace.

Someone, somebody, no other, none else, an individual.

*V.* To be one, to be alone &c. ; to isolate, insulate, set apart.

*Adj.* One, sole, single, individual, apart, alone, lone, isolated, solitary, lonely, lonesome, desolate, dreary, insular, insulated, disparate, discrete, detached (44).

Unaccompanied, unattended; *solus*, single handed; singular, odd, unique, unpeated, axygous.

Insecable, inseverable, irresolvable, indiscerptible, compact.

*Adv.* Singly, &c., alone, by itself, per se, only, apart, in the singular number, in the abstract; one by one. One and a half, sesqui-.

**Dictionnaire idéologique (1859)**

**87 UNITÉ** *N.* individualité, solitude, isolation, isolement (893), abstraction.

Un, unité, as, monôme, monolithe. Quelqu'un, individu.

*V.* Être un, être seul ; *adj.* individualiser, isoler etc., mettre à part, mettre à l'écart.

*Adj.* Un, seul, unique, individuel, à part, isolé, esseulé, solitaire, détaché (44),

Solus, singulier, impair.

Insécable, compacte.

*Adv.* Seulement, isolément, *per se*, à part, à l'écart, un à un, tout en gros.

Un et demi, *sesqui*. [...]

Le parallélisme structurel entre le *Thesaurus of English Words and Phrases* et le *Dictionnaire idéologique* se poursuit donc au niveau des entrées. Les informations lexicales encodées dans le thésaurus anglais servent à Théodore Robertson de cadre à la conception de ses articles de dictionnaire. La structure d'un article repose sur une suite, préétablie par Roget, des mots et des locutions anglaises, attachée à une notion conceptuelle. L'article 87. *Unité* du *Dictionnaire idéologique* présente par exemple la suite *individualité - solitude - isolation*. Elle reflète le découpage conceptuel qu'opère l'article 84. *Unity* du *Thesaurus of English Words and Phrases* par la suite : *individuality - solitude - isolation*. L'exemple est emblématique pour la constitution des entrées lexicales au sein du *Dictionnaire idéologique*.

Le *Dictionnaire idéologique* constitue - au niveau des catégories conceptuelles et au niveau des entrées lexicales - la première transposition du *Plan of classification* vers une autre langue européenne. Le réaménagement de quelques articles est, dans cette perspective,

<sup>62</sup> Tableau 4 de S. Reusch, élaboré d'après les informations trouvés dans Hüllen (2009), pp. 86-89, dans le *Dictionnaire idéologique* (1859), article 87. Unité, p. 30 et dans Roget's *Thesaurus* (1856), article 87. Unity, p. 27.

d'un grand intérêt. Il montre que le *Dictionnaire idéologique* n'est pas, en dépit des ressemblances relevées, une traduction littérale irréfléchie du *Thesaurus of English Words and Phrases*. Il existe deux circonstances dans lesquelles Théodore Robertson se détache de son projet d'une imitation exacte du *Plan of classification*, au niveau des articles.<sup>63</sup>

1. Dans le premier cas, la popularité de certains mots français l'emporte sur une traduction exacte. Dans ce cas, Robertson opte pour une nouvelle désignation des concepts en question:

**Tableau 5: Les concepts redésignés par Robertson<sup>64</sup>**

Roget (1852)			Robertson (1859)	
110 <i>Diuturnity</i>	111 <i>Transientness</i>	→	110 <i>Longue Durée</i>	111 <i>Courte Durée</i>
518 <i>Intelligibility</i>	519 <i>Unintelligibility</i>	→	518 <i>Clarté</i>	519 <i>Obscurité</i>
914 <i>Pity</i>	914a <i>Pitilessness</i>	→	914 <i>Pitié</i>	914 bis <i>Inhumanité</i>

2. Dans le second cas, la disposition des articles en paires d'opposés montre une incohérence, voire une erreur. Robertson opte pour une restructuration des articles en question<sup>65</sup> :

**Tableau 6 : Les articles restructurés par Robertson<sup>66</sup>**

Roget (1852)			Robertson (1859)	
140 <i>Change</i>	140a <i>Abscence of change</i>		140 <i>Changement</i>	141 <i>Permanence</i>
141 <i>Cessation</i>	142 <i>Permanence</i>		142 <i>Cessation</i>	143 <i>Continuation</i>
	143 <i>Continuance</i>			
987 <i>Piety</i>	988 <i>Irreligion</i>		987 <i>Piété</i>	988 <i>Impiété</i>
	989 <i>Impiety</i>			989 <i>Irréligion</i>

Le *Dictionnaire idéologique* n'est pas - comme le pourrait faire songer la description de l'*approche expansive* du chapitre précédent - un descendant historique du *Thesaurus of English Words and Phrases*. Les deux ouvrages, publiés avec un intervalle de seulement sept années, influencent leur conception respectivement. L'échange intellectuel étroit entre

<sup>63</sup> Cf. Hüllen (2009), pp.85-87.

<sup>64</sup> Tableau 5 : Adopté littéralement de Hüllen (2009), p.86, corrigé par S. Reusch ( 914 *Piety* → 914 *Pity* ).

<sup>65</sup> Cf. Hüllen (2009), pp.85-87.

<sup>66</sup> Tableau 6 : Adopté littéralement de Hüllen (2009), p.86.

les deux auteurs, Roget et Robertson, en vue d'une *synchronisation des ouvrages*, fait que les réaménagements français sont aussi discutés pour une nouvelle publication du *Thesaurus of English Words and Phrases* en Angleterre. La réédition de 1899 corrige les arrangements critiquables d'après la traduction de Théodore Robertson <sup>67</sup> :

**Tableau 7 : Correction des arrangements critiquables par Roget** <sup>68</sup>

Robertson (1859)		Roget (1911) <sup>69</sup>	
140 <i>Changement</i>	141 <i>Permanence</i>	140 <i>Change</i>	141 <i>Permanence</i>
142 <i>Cessation</i>	143 <i>Continuation</i>	142 <i>Cessation</i>	143 <i>Continuance</i>
987 <i>Piété</i>	988 <i>Impiété</i>	987 <i>Piety</i>	988 <i>Impiety</i>
	989 <i>Irréligion</i>		989 <i>Irreligion</i>

Théodore Robertson constitue le *Dictionnaire idéologique* sur le modèle lexicographique du *Thesaurus of English Words and Phrases*, car il est profondément convaincu que :

« Le *plan de classification* adopté dans cet ouvrage étant applicable à toutes les langues [...] pourrait composer de la même manière un dictionnaire allemand, italien, espagnol [...]. » <sup>70</sup>

Le *Plan de classification* est donc pour lui, et pour Peter Mark Roget, une *structure* qui permet d'aligner plusieurs langues européennes - en vue de leur classement d'idées - au sein d'un *lexique polyglotte*. Cet alignement présente pour lui une première approche de la création d'une langue universelle <sup>71</sup>, car :

« Les métaphysiciens, livrés à de profondes recherches sur la philosophie du langage, trouveront une assistance matérielle, dans un travail qui leur aura déblayé le terrain, par une analyse préalable et par la classification des idées ; car cette classification est la véritable base de celle des mots qui sont les symboles des idées. » <sup>72</sup>

<sup>67</sup> Cf. Hüllen (2009), p. 86.

Tableau de S. Reusch, adapté d'après un tableau proposé par Hüllen (2009), p.86.

<sup>68</sup> Cf. Hüllen (2009), p. 86.

<sup>69</sup> Tableau 7 de S. Reusch, élaboré d'après les informations trouvés dans le *Dictionnaire idéologique* (1859), Tableaux synoptiques des catégories, p. xvi-xxvii et dans Roget's *Thesaurus* (2002), A selection of pages to the first édition : Plan of Classification, p. xxxiv.

<sup>70</sup> Cité d'après *Dictionnaire idéologique* (1859), Introduction, p. xii.

<sup>71</sup> Cf. Quemada (1967), p. 381.

<sup>72</sup> Cité d'après *Dictionnaire idéologique* (1859), Introduction, p. xiii.

L'alignement de ces *symboles des idées* - de ces *concepts* figurant au *Thesaurus of English Words and Phrases* et au *Dictionnaire idéologique* - en vue de la création d'une langue universelle, contribue, selon les auteurs, à une entente des peuples, car :

« Rien ne pourrait amener plus directement l'union et l'harmonie parmi toutes les races humaines que le renversement de cette barrière, - la diversité des langues, - qui s'oppose maintenant à l'échange de la pensée et du bon vouloir entre l'homme et son semblable. »<sup>73</sup>

Malgré cette *vision flamboyante* des deux auteurs, Roget et Robertson, la création d'un *lexique polyglotte* sur le modèle lexicographique du *Thesaurus of English Words and Phrases* ne fut pas couronnée de succès. Le projet lexicographique va rester prisonnier de l'illusion du temps, qui ignore le fait que chaque langue opère un découpage conceptuel différent de la réalité.

### 1.2.3 Thésaurus Larousse : Une parenté exprimée par *similarité organisationnelle*

Le *Thésaurus Larousse* (1992) est une adaptation moderne du modèle lexicographique du *Thesaurus of English Words and Phrases*. À la recherche d'une nouvelle approche du classement analogique, l'éditeur Larousse s'intéresse au « [...] projet de Roget de prendre la langue dans un réseau conceptuel maillé couvrant méthodiquement l'ensemble des champs notionnels possibles [...] »<sup>74</sup>. Les lexicographes s'intéressent à la « vision globale »<sup>75</sup> du classement par centres d'intérêt, mais ne veulent pas s'attacher à la conceptualisation inhérente de l'ouvrage. Daniel Péchoin, le directeur éditorial du *Thésaurus Larousse*, explique son approche au thésaurus anglais comme suivant :

« L'on sait qu'il n'existe pas de coïncidence parfaite entre les significations des mots de deux langues et qu'a fortiori les champs notionnelles de l'une et de l'autre ne sont pas directement superposables. [...] Il n'en reste pas moins qu'un *Thésaurus* français devait opérer un découpage notablement différent de celui de son devancier avait utilisé pour l'anglais dans l'univers des significations de notre langue. »<sup>76</sup>

Le *Thésaurus Larousse* se construit, par conséquent, dans une *approche par dérivation* par rapport au *Thesaurus of English Words and Phrases*. Une comparaison détaillée des deux

---

<sup>73</sup> Cité d'après *Dictionnaire idéologique* (1859), Introduction, p. xiii.

<sup>74</sup> Cité d'après *Thésaurus Larousse* (1992), Préface, p.vii.

<sup>75</sup> Cité d'après Péchoin (2001), p. 117.

<sup>76</sup> Cité d'après *Thésaurus Larousse* (1992), Préface, pp.vi/vii.

ouvrages montre que leur structure classificatoire est *divergente*: Le ‘*Plan of classification*’ et le ‘*Plan de classification*’ sont différents (cf. tableau 8).<sup>77</sup>

**Tableau 8 : Comparaison de la structure classificatoire entre Roget et Péchoin**

<u>Roget (1852)</u>			<u>Péchoin (1991)</u>		
<b>Plan of classification</b>			<b>Plan de classification</b>		
			<b>A. LE MONDE</b>		
<b>I. Abstract relations</b>	1. Existence	2. Relation	<b>I. Grandes notions</b>	1. Existence	2. Relation
	3. Quantity	4. Order		3. Cause	4. Ordre
	5. Number	6. Time		5. Quantité	6. Nombre
	7. Change	8. Causation			
<b>II. Space</b>	1. Generally	2. Form	<b>II. Espace</b>	1. Dimensions	2. Contours
	3. Dimensions	4. Motion		3. Forme	4. Structure
				5. Position	6. Localisation
<b>III. Matter</b>	1. Generally	2. Inorganic	<b>III. Temps</b>	1. Généralités	2. Chronologie
	3. Organic			3. Évolution	
<b>IV. Intellect</b>	1. Formation of ideas	2. Communication of ideas	<b>IV. Mouvement</b>	1. Mouvement	2. Force
<b>V. Volition</b>	1. Individual	2. Intersocial	<b>V. Matière</b>	1. Sciences de la matière	2. Propriétés de la matière
				3. Éléments et matériaux	4. L’environnement terrestre
<b>VI. Affections</b>	1. Generally	2. Personal	<b>VI. Vie</b>	1. Généralités	2. Plantes
	3. Sympathetic	4. Moral		3. Animaux	
	5. Religious				<i>etc.</i>

Les deux ouvrages présentent deux systèmes de classement qui diffèrent dans la dénomination des catégories qu’ils emploient et dans la séquence dans laquelle elles sont rendues. La convergence entre les tableaux classificatoires est de nature *pragmatique*. Dans la première étape de la conception du *Thésaurus Larousse*, les lexicographes s’appuient d’abord sur le modèle classificatoire de Peter Mark Roget. Leur projet lexicographique commence par la traduction du *Plan of classification* - qui sert de grille

<sup>77</sup> Tableau 8 de S. Reusch, élaboré d’après les informations trouvés dans Roget’s Thesaurus (2002), A selection of pages to the first édition : Plan of Classification, p. xxxiv et dans le Thésaurus Larousse (1992), Sommaire, pp.xv-xxi.



pour la distribution du travail. Après la rédaction des premiers articles, la *grille* se divise entre les collaborateurs qui, à leur tour, réaménagent le contenu du *Thesaurus of English Words and Phrases* en vue d'un classement lexical français. Le plan de classification du *Thésaurus Larousse* reflète cette démarche (cf. tableau 9). La première catégorie I. *Grandes notions* adopte six classes de la catégorie I. *Abstract relations* du *Thesaurus of English Words and Phrases*. Deux classes (6. *Time* et 7. *Change*) ne sont pas transposées dans cette catégorie. La première (6. *Time*), sert de descripteur pour une nouvelle catégorie créée : III. *Temps*, la seconde (7. *Change*) ne figure plus au système de catégories : elle présente un article qui figure *dans* la catégorie III. *Temps*.

**Tableau 9 : La transposition du « Plan of classification » au Thésaurus Larousse** <sup>78</sup>

<b>I. Abstract relations</b>	1. Existence	2. Relation	<b>I. Grandes notions</b>	1. Existence	2. Relation
	3. Quantity	4. Order		3. Cause	4. Ordre
	5. Number	<u>6. Time</u>		5. Quantité	6. Nombre
	<u>7. Change</u>	8. Causation			[...]
				1. Généralités	2. Chronologie
			<b>III. Temps</b>	3. Évolution	
				193. Changement	

Le schéma montre que l'approche dérivationnelle tend à reproduire deux thésauri structurellement divergents. Une comparaison subséquente des deux ouvrages en vue de leur *convergence* est uniquement possible, lorsque le '*Plan of classification*' du modèle anglais est repérable comme *structure sous-jacente* de la classification. Le *Thésaurus Larousse* organise ainsi les concepts numéraux de la même manière que le *Thesaurus of English Words and Phrases*, comme le montre un schéma comparatif :

<sup>78</sup> Tableau 9 de S. Reusch, élaboré d'après les informations trouvés dans le *Thésaurus Larousse* (1992), Sommaire, pp. xv/xvi et dans Roget's *Thesaurus* (1987), Tabular synopsis of categories, pp. xlv/xlv

**Tableau 10 : Similarité organisationnelle au niveau de la macrostructure** <sup>79</sup>

**Thesaurus of English Words (1987)**

**Thésaurus Larousse (1992)**

**CLASS I : ABSTRACT RELATIONS**

**CLASSE I : GRANDES NOTIONS**

**V. Number**

**VI. Nombre**

1. Abstract

100. Nombre

85. Number

101. Zéro

...

2. Determinate

**88. Unity**

89. Accompani-  
{ -ment

**90. Duality**

91. Duplication

92. Bisection

**93. Triality**

94. Triplication

95. Trisection

**96. Quaternity**

...

107. Infinity

**102. Un**

**103. Deux**

**104. Trois**

**105. Quatre**

...

113. Cent

114. Mille

115. Infini

Au niveau des entrées lexicales, on ne peut pas parler d'une *convergence* entre les deux ouvrages. Les concepts anglais (comme *unity*) servent dans le *Thésaurus Larousse* uniquement comme *point d'ancrage* pour la conception des nouveaux articles, qui suivent une logique française. Une comparaison des articles 87.*Unity* et 102.*Un* montre que leur convergence lexicale (indiquée par un soulignement) est très faible :

<sup>79</sup> Tableau 10 de S. Reusch, élaboré d'après les informations trouvés dans le *Thésaurus Larousse* (1992), Sommaire, pp. xv/xvi et dans Roget's *Thesaurus* (1987), Tabular synopsis of categories, pp. xlv/xlv

**Tableau 11 : Similarité organisationnelle au niveau de la microstructure**<sup>80</sup>

**Thesaurus of English Words (1987)**

**Thésaurus Larousse (1992)**

**88 UNITY.** *N.* unity; oneness.; individuality; singleness, solitariness, solitude 893; isolation, abstraction.

One, unit, ace.

Someone, somebody, no other, none else, an individual.

*V.* To be one, to be alone &c. ; to isolate, insulate, set apart.

*Adj.* One, sole, single, individual, apart, alone, lone, isolated, solitary, lonel, lonesome, desolate, dreary, insular, insulated, disparate, discrete, detached (44).

Unaccompanied, unattended; *solus*, single handed; singular, odd, *unique*, unrepeatd, axygous.

Insecable, inseverable, irresolvable, indiscerptible, compact.

*Adv.* Singly, &c., alone, by itself, *per se*, only, apart, in the singular number, in the abstract; one by one. One and a half, *sesqui*-.

**102 UN** *N.* 1. un ; unité - L'Un. [philos.]

2. Un à un, un par un. - L'un après l'autre. - L'un, l'une ; les uns, les unes ; l'un l'autre. - Comme pas un. - Tant l'un (*deux francs l'un*). Tant l'unité, tant la pièce.

3. Un ; un seul. - Un pour tous, tous pour un 23.3 ; « Et s'il n'en reste qu'un, je serai celui-là » (V. Hugo).

4. jeux. - Un (*le un*) ; as - Manillon ; anc. : ambesas, baste, besas, bezet, spadille. - Parier à dix, cent, *n* contre un.

5. Singleton [math.]

6. Premier (*le premier*). - Primidi [hist.] 176.

7. sc. éduc. - Primaire (*le primaire*) ; première (*la première*) ; première supérieure. - Première ou première classe [transp.]. - Premier de l'an.

8. Unicité. - Unicisme ; uniciste (*un uniciste*). - Uniformité 24.

*V.* 9. Unifier ; unir. - Uniformiser.

*Adj.* 10. Un. - Une fois ; pour une fois, pour un coup [fam.] [...]

13. Unique ; inégal ; sans équivalent, sans pareil ; à nul autre pareil. [...]

*Adv.* 15. Uniquement, seulement.

16. Premièrement. - Primo [lat.]. Unièment (*en vingt, trente, etc., et unième position*).

17. D'abord ; en premier [fam.]

*Int.* 18. Et d'un !

*Aff.* 19. Uni- ; mono- ; primo-.

Au niveau des informations lexicales encodées, la divergence entre la conception du *Dictionnaire idéologique* et du *Thésaurus Larousse* s'aperçoit clairement. La comparaison des articles (88. *Unity* et 102. *Un*) montre à quel point la prise en compte des expressions françaises change le caractère de l'article source. En référence à cet exemple, on peut dire que l'innovation et la réception du modèle lexicographique du *Thesaurus of English Words*

<sup>80</sup> Tableau 11 de S. Reusch, élaboré d'après les informations trouvées dans le *Thésaurus Larousse* (1992), article 102. Un, p. 60 et dans Roget's *Thesaurus* (1987), 88. *Unity*, pp. 54/55.

*and Phrases* par les lexicographes modernes s'inscrit sur le plan des *structures pragmatiques*, qui indiquent que le *Thésaurus Larousse* s'oriente à la langue cible.<sup>81</sup>

Sur le plan historique, le *Thésaurus Larousse* est le prolongement d'un projet lexicographique entamé par le *Dictionnaire idéologique* : celui de la création d'un thésaurus français sur le système de classement du *Thesaurus of English Words and Phrases*. Sur le plan théorique, le *Thésaurus Larousse* est le précurseur d'un projet lexicographique encore à entamer : celui de la création d'un *nouveau* thésaurus, qui répondre aux exigences actuelles de la langue française. Dans cette perspective, la compréhension du *système de classement* de ce type de thésaurus est importante.

### 1.3 Structure et fonctionnement du thésaurus français

#### 1.3.1 Système de classement

*Un thésaurus constitue un dictionnaire hiérarchisé. Le vocabulaire est structuré selon des termes génériques (ou descripteurs), représentant chacun un concept. Les termes sont arrangés dans une hiérarchie de concepts, destinée à la classification et à l'organisation de concepts par matière. Ils sont liés par un ensemble de relations exprimant les rapports sémantiques entre les concepts.*<sup>82</sup>

Cette définition unifie deux fonctions principales d'un thésaurus, à savoir la fonction de classification des concepts et la fonction d'organisation des concepts.<sup>83</sup> En tant que système de classification un thésaurus attribue, de façon univoque et réciproque, un descripteur à un concept.<sup>84</sup> En tant que système d'organisation un thésaurus tend à l'enregistrement des relations qui s'établissent entre les concepts classés. Un thésaurus se compose donc d'un *ensemble de termes* et d'un *ensemble de relations* entre ces descripteurs.<sup>85</sup> Les relations entre les termes servent à obtenir une association univoque et réciproque entre le concept, son *extension* et son *intension*. L'*extension* d'un concept désigne l'*objet*, donc la classe d'éléments ou de choses désignées par le descripteur. L'*intension* d'un concept indique le *sens*, donc l'ensemble des attributs ou des propriétés

---

<sup>81</sup> Cf. Hüllen (2009), p. 69.

<sup>82</sup> Définition de S. Reusch, récite de Bianucci (1992), p.60 et Taddei Elmi (1981), p. 847 et Reisinger (1977), p.140 et Renard et al. (2009), p.88 et de Carlisle (2011), p. 2.

<sup>83</sup> Cf. Taddei Elmi (1981), p. 847 ; cf. Bianucci (1992), p. 62.

<sup>84</sup> Un concept est considéré dans ce cas comme une « signification normalisée », cf. Bianucci (1992), p. 62.

<sup>85</sup> Cf. Taddei Elmi (1981), p. 847.

qui caractérisent chaque objet désigné par le descripteur (cf. **chapitre 3.1**).<sup>86</sup> Le classement du thésaurus montre l'*extension* et l'*intension* d'un concept. Les *termes* sont susceptibles de prendre un sens spécifique suivant leur regroupement en domaines thématiques et leur organisation en relations sémantiques. La structure d'un thésaurus aide ainsi à la désambiguïsation de concepts et permet un repérage facile de leur *signification*.<sup>87</sup>

### 1.3.2 Nature des termes

*Un terme est un signe linguistique qui est constitué d'une désignation renvoyant à un concept. La désignation est de l'ordre de la langue et le concept de l'ordre de la pensée.*<sup>88</sup>

Un thésaurus présente deux types de termes. La hiérarchie consiste en *termes préférés* (descripteurs). Ils représentent sans ambiguïté des concepts. Le classement et la recherche se fait à l'aide de ces termes.<sup>89</sup> Pour un repérage facile, ces termes ont la forme d'un nom ou d'un syntagme nominal. Les articles consistent en *termes non-préférentiels* (non-descripteurs) qui sont des synonymes, des quasi-synonymes ou des variantes orthographiques d'un descripteur.<sup>90</sup> Ils constituent l'ensemble des informations qui suivent l'entrée et fournissent une vision globale sur ce qu'est le concept.<sup>91</sup> La relation entre *descripteur* et *non-descripteur* est dynamique et s'exprime par des renvois réciproques.<sup>92</sup>

Le Thésaurus Larousse (1992) distingue trois types de descripteurs : Les termes génériques, les termes spécifiques et les termes associés.

#### *Terme générique (TG)*

Le *terme générique* est un descripteur qui désigne un concept général auquel sont subordonnés des *termes spécifiques* dans une hiérarchie. Le concept général englobe donc les notions plus fines qui sont introduites par les *termes spécifiques*.<sup>93</sup>

Ex. 799. Outils

---

<sup>86</sup> Cf. Depecker (1999), pp. 42/43 ; cf. Wille (2005), p.2 ; cf. Taddei Elmi (1981), pp. 847/848 ; cf. Rey (1976), pp. 84-86.

<sup>87</sup> Cf. Bianucci (1992), pp. 62/63.

<sup>88</sup> Cf. Depecker (1999), p. 138 ; cf. Hodgson (1997), pp. 44/45 ; Cf. Dictionnaire de linguistique (1994), Article : Terme, p. 480.

<sup>89</sup> Cf. Carlisle (2011), Guidance Manual, p. 6 ; cf. Carlisle (2011), Glossary, p. 5 ; cf. Gouvernement du Québec (2011), Politique d'édition du TAG, pp. 2/3.

<sup>90</sup> Cf. Carlisle (2011), Glossary, p. 5 ; cf. Gouvernement du Québec (2011), Politique d'édition du TAG, p. 3.

<sup>91</sup> Cf. Rey-Debove (1971), p.151 ; cf. Gouvernement du Québec (2011), Politique d'édition du TAG, p. 3.

<sup>92</sup> Cf. Rey-Debove (1989), pp. 931-937.

<sup>93</sup> Cf. Carlisle (2011), Glossary, p. 4. Exemples tirés du Thésaurus Larousse (1992), article 799. Outils, pp. 555/556.

TG : *Outil*

TS : *Marteau, Pince, Lime etc.*

### *Terme spécifique (TS)*

Le *terme spécifique* est un descripteur qui désigne une notion restreinte qui est incluse dans un concept général auquel il se réfère. Il peut être subordonné à un ou plusieurs *termes génériques* dans une hiérarchie.<sup>94</sup>

Ex. 848. Maison

TS : *Villa*

TG : *Maison*

### *Terme associé (TA)*

Le *terme associé* est un descripteur que l'on associe à un autre descripteur de façon conceptuelle et non hiérarchique. La relation d'association établie indique les liens de signification entre les deux termes.<sup>95</sup>

Ex. 286. Arbres

TG : *Arbre*

TA : *Déforestation* (812. Arboriculture)

## **1.3.3 Nature des relations**

La structure sémantique d'un thésaurus comprend quatre types de relations : La relation hiérarchique, la relation d'équivalence, la relation associative et la relation d'opposition. Ces relations sémantiques permettent de situer un *terme* dans un environnement conceptuel.<sup>96</sup>

### *1. Relation hiérarchique*

La relation hiérarchique est le lien principal entre un *terme générique* et ses *termes spécifiques*. Elle met en évidence les relations « genre/espèce » et « tout/partie » qui existent entre les entités donnés.<sup>97</sup> Par exemple, le terme générique *maison* représente un tout (ou une classe) et ses termes spécifiques tels que *escalier* ou *fenêtre* renvoient à ses parties (ou ses éléments). Le terme *maison* représente aussi un genre, dont les termes

---

<sup>94</sup> Cf. Carlisle (2011), Glossary, p. 5. Exemples tirés du Thésaurus Larousse (1992), article 848. Maison, pp. 600-602.

<sup>95</sup> Cf. Carlisle (2011), Glossary, p. 5. Exemples tirés du Thésaurus Larousse (1992), articles 286. Arbres (p. 175) et 812. Arboriculture (p. 568).

<sup>96</sup> Cf. Gouvernement du Canada (2009), *Thésaurus des sujets de base du gouvernement du Canada*, consultable sur : [www.thesaurus.gc.ca](http://www.thesaurus.gc.ca)

<sup>97</sup> Cf. Hodgson (1997), p. 45. ; cf. Gouvernement du Québec (2011), Politique d'édition du TAG, p. 6.

spécifiques tels que *maison de campagne* ou *villa* précisent l'espèce (le type de maison).<sup>98</sup> Dans le sens vertical les relations hiérarchiques constituent le cadre principal d'un thésaurus. Elles permettent de le représenter en forme d'arbre, construit sur une chaîne conceptuelle allant du terme le plus général au terme le plus spécifique.<sup>99</sup>

## 2. Relation d'équivalence

La relation d'équivalence est un lien qui s'instaure - au sein d'un article - entre deux ou plusieurs *termes spécifiques* qui ont la valeur de synonyme ou de quasi-synonyme. Ils forment un ensemble qui se réfère à un concept (ou *terme générique*).<sup>100</sup> La synonymie s'introduit dans un thésaurus par différentes descriptions possibles d'un même concept. Le concept peut être décrit par des différents niveaux de spécialisation, des registres de langue et des variations géographiques.<sup>101</sup> La synonymie sert donc par sa nature à enrichir le contenu lexical du dictionnaire.<sup>102</sup>

## 3. Relation associative

La relation associative exprime le rapport qui existe entre deux termes qui ont une proximité sémantique sans appartenir à un même article.<sup>103</sup> Les termes ne sont pas liés par des relations de hiérarchie ou d'équivalence, mais par une relation d'analogie. La relation d'analogie fournit un enrichissement sémantique, en guidant l'utilisateur sur les sujets connexes au-delà de l'article.<sup>104</sup> Ainsi, peuvent être liés des termes tels que '*arbre*' et '*déforestation*'.<sup>105</sup>

## 4. Relation d'opposition

La relation d'opposition est un lien qui s'instaure - au sein de la hiérarchie des notions - entre deux *termes génériques* qui ont - par leur communauté de sens - une valeur d'antonymes.<sup>106</sup> L'antonymie définit surtout des termes qui représentent des qualités

---

<sup>98</sup> Exemples tirés du Thésaurus Larousse (1992), article 848. Maison, pp. 600-602.

<sup>99</sup> Cf. Gouvernement du Québec (2011), Politique d'édition du TAG, pp. 8/9 .

<sup>100</sup> Cf. Gouvernement du Québec (2011), Politique d'édition du TAG, pp. 8/9 ; cf. Carlisle (2011), Guidance Manual, p. 7 ; cf. Old (2003), pp. 12-16.

<sup>101</sup> Cf. Depecker (1999), p.131.

<sup>102</sup> Cf. Lamizet (1975), p. 18 ; cf. Old (2003), pp. 12-16.

<sup>103</sup> Cf. Hodgson (1997), p. 45 ; cf. Carlisle (2011), Guidance Manual, p. 7 ; cf. Gouvernement du Québec (2011), Politique d'édition du TAG, p. 7.

<sup>104</sup> Cf. Bianucci (1992), p. 78 ; cf. Niklas-Salminen (1997), p.113 ; cf. Cf. Hodgson (1997), p. 45.

<sup>105</sup> Cf. Thésaurus Larousse (1992), articles 286. Arbres (p. 175) et 812. Arboriculture (p. 568).

<sup>106</sup> Cf. Bianucci (1992), p. 74 ; cf. Old (2003), p. 16 et p. 22 ; cf. Gouvernement du Québec (2011), Politique d'édition du TAG, p. 9.

(*beauté - laideur*), des quantités (*richesse - pauvreté*), des dimensions (*grandeur - petitesse*) ou des rapports chronologiques (*passé - futur*).<sup>107</sup> Pour le classement sémantique du thésaurus, l'antonymie est logiquement indispensable, car elle reflète une tendance générale de l'homme de vouloir catégoriser son expérience en termes de contrastes dichotomes.<sup>108</sup>

### 1.3.4 Unités de classement

Trois grandes unités de classement servent dans un thésaurus à la présentation de l'ensemble des termes : Le *Plan de classification*, le *corpus* et l'*index*. Le *Plan de classification* et le *corpus* organisent les termes selon les idées qu'ils expriment. L'*index* classe les termes selon les variantes linguistiques qu'ils représentent.<sup>109</sup>

#### 1) Le Plan de classification

Le *Plan de classification* fournit une vue synthétique sur le système de classement du thésaurus.

Le Thésaurus Larousse repose sur trois grandes catégories : **A. Le Monde**, **B. L'Homme** et **C. La Société**. Chaque ensemble classificatoire est lui-même l'objet des nombreux sous-ensembles, ce qui crée une « hiérarchie des notions ».<sup>110</sup> Si l'on consulte la catégorie **A. Le Monde** on aperçoit qu'elle comprend six sous-catégories **I. Les Grandes Notions**, **II. L'Espace**, **III. Le Temps**, **IV. Le Mouvement**, **V. La Matière** et **VI. La Vie**. La matière de la classe **I. Les Grandes Notions** est distribuée sous six divisions **1. Existence**, **2. Relation**, **3. Cause**, **4. Ordre**, **5. Quantité** et **6. Nombre**.<sup>111</sup> Dans chaque division se trouve une liste des descripteurs choisis pour représenter les concepts sous lesquels on range les informations lexicales, en forme d'article de dictionnaire.<sup>112</sup> La figure 3 montre le système de classification du Thésaurus Larousse.

---

<sup>107</sup> Cf. Niklas-Salminen (1997), p.114.

<sup>108</sup> Cf. Niklas-Salminen (1997), p.114.

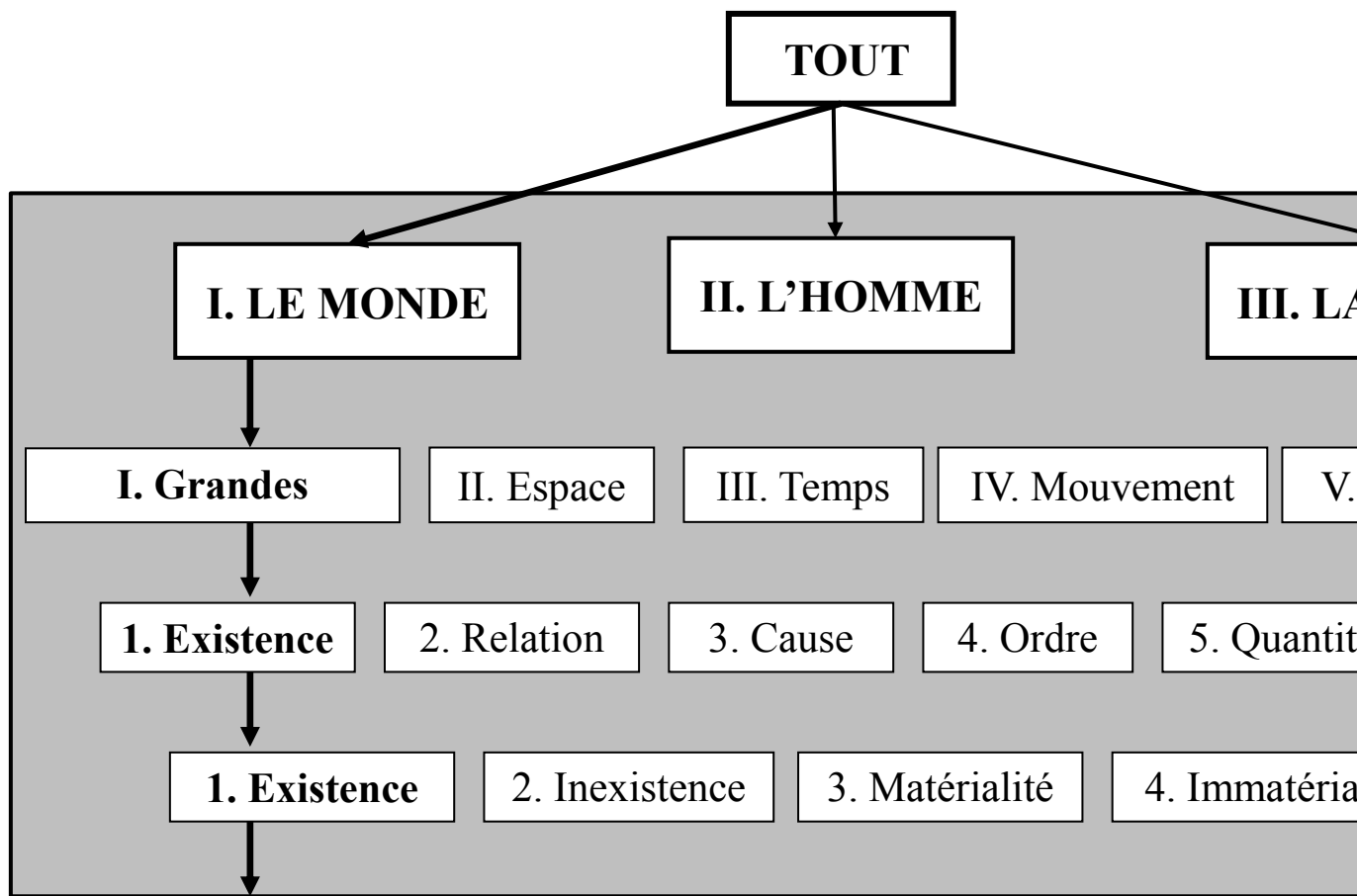
<sup>109</sup> Cf. Boulanger (1993), p.134.

<sup>110</sup> Cf. Boulanger (1993), p.132.

<sup>111</sup> Cf. Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.

<sup>112</sup> Cf. Boulanger (1993), p.132.





## ARTICLE 1. Existence

2 Métaphysique 478.2, ontologie; « il y a une science (*la métaphysique*) en tant qu'être et les attributs qui lui appartiennent essentiellement. *Cogito ergo sum.* (lat. « je pense donc je suis. » Descartes), [...].

3 Actualité, réalité (*la réalité*) – Concret (*le concret*), réel (*le réel*),

4 Individu, personne 307, personnalité, sujet. – Âme, conscience, esprit.

5 Abstraction, entité, noumène. – Concept, **idée**, imagination (*une idée*).

6 Création, réalisation; créature. – Origine, souche. – Naissance 313.

7 Jugement d'existence (opposé à jugement de valeur). \_ Verbe d'existence (copule) [gramm.] – Existence juridique [dr.]

V 8 **Exister, être**; vivre. – « Être ou ne pas être, voilà la question. » (to be, that is the question, Shakespeare, Hamlet). – Régner, [...].

**Adj.** 12 Existential; **existentiel**, ontique, ontologique; essentiel [...].

**Adv.** 14 Existentiellement, ontologiquement; essentiellement. – Actuellement.

**Aff.** 15 Onto-.

### Figure 3 : Le système de classement du Thésaurus Larousse <sup>113</sup>

Dans un thésaurus, les articles sont disposés aussi souvent que possible par paires des notions opposées ou corrélatives : *existence* est opposé à *inexistence*, *matérialité* est opposé à *immatérialité* (cf. figure 3). <sup>114</sup> Dans le corpus, les articles des notions opposées figurent sur une même page ou double-page. Selon les auteurs, « cette disposition permet d’embrasser d’un seul coup d’œil l’ensemble des synonymes, des contraires et des analogies d’un mot donné. ». <sup>115</sup>

#### 2) Le corpus

Au niveau des articles, le *corpus* recense les *termes spécifiques*, les mots et locutions relevant des différentes notions conceptuelles.

Dans un thésaurus, les articles sont ordonnés selon un double principe, grammatical et sémantique. Les catégories grammaticales déterminent, selon un ordre convenu, la succession des paragraphes en noms, verbes, adjectifs, adverbes, conjonctions, prépositions et phrases. <sup>116</sup> Figure 4 montre l’article 1. Existence du Thésaurus Larousse.

#### ARTICLE 1. Existence

- N** 1 **Existence**, être (*l'être; l'Être*) 540, présence, réalité; essence, substance [...]  
2 Métaphysique 478.2, ontologie; « il y une science (*la métaphysique*) qui étudie l'Être en tant qu'être et les attributs qui lui appartiennent essentiellement. » (Aristote) . – *Cogito ergo sum*. (lat. « je pense donc je suis. » Descartes), [...].  
3 Actualité, réalité (*la réalité*) – Concret (*le concret*), réel (*le réel*), présent (le présent).  
4 Individu, personne 307, personnalité, sujet. – Âme, conscience, esprit, [...].  
5 Abstraction, entité, noumène. – Concept, **idée**, imagination (*une imagination*) 404 [...]  
6 Création, réalisation; créature. – Origine, souche. – Naissance 313, génération [...]  
7 Jugement d'existence (opposé à jugement de valeur). \_ Verbe d'existence (opposé à copule) [gramm.] – Existence juridique [dr.]  
**V** 8 **Exister, être**; vivre. – « Être ou ne pas être, voilà la question. » (trad. de *to be or not to be, that is the question*, Shakespeare, Hamlet). – Régner, [...].  
**Adj.** 12 Existential; **existentiel**, ontique, ontologique; essentiel [...].  
**Adv.** 14 Existentiellement, ontologiquement; essentiellement. – Actuellement [...].  
**Aff.** 15 Onto-.

<sup>113</sup> Figure 3 de S. Reusch, élaboré d'après les informations trouvés dans le Sommaire (pp. vx-xxi) et dans le Corpus (p. 1, article 1. Existence) du Thésaurus Larousse (1992).

<sup>114</sup> Cf. Thésaurus Larousse (1992), Corpus, pp. 1-3.

<sup>115</sup> Cité d'après Thésaurus Larousse (1992), Préface, p.vi.

<sup>116</sup> Cf. Péchoin (2001), p.91.

#### Figure 4 : L'article 1. Existence du Thésaurus Larousse <sup>117</sup>

L'article **1. Existence** présente, par exemple, une catégorie des verbes, parmi lesquels figurent *exister*, *être* et *vivre* (cf. figure 4). À l'intérieur de chaque catégorie grammaticale, les termes spécifiques sont organisés en séries sémantiques cohérents qui forment des paragraphes numérotés. <sup>118</sup> Dans l'article **1. Existence** on distingue sept séries sémantiques des noms (cf. figure 4). La première contient la notion de base, exprimée par les noms *existence*, *être* et *réalité*, la seconde la notion scientifique, exprimée par les noms *métaphysique* et *ontologique*, etc. Au sein de chaque série sémantique, un appareil diacritique facilite le décodage des nuances de sens : Le *tiret* indique une division sémantique forte, le *point-virgule* des nuances sémantiques plus fines. <sup>119</sup> Un autre code microstructurel complète la physionomie de l'article : Le *renvoi numéroté*. Il encourage l'utilisateur à poursuivre la recherche d'informations dans le réseau sémantique du thésaurus, en établissant des interconnexions entre les différentes notions conceptuelles (des articles). <sup>120</sup>

#### 3) L'index alphabétique

La navigation dans le *corpus* est facilitée par la présence d'un *index alphabétique*, qui répertorie la totalité des *items* attachés aux articles. <sup>121</sup>

La figure 5 montre un extrait d'index du Thésaurus Larousse pour l'item *existence*. <sup>122</sup> L'item ne renvoie pas uniquement à l'article de base (1. Existence), mais aussi aux articles de sens voisin (310. Vie ; 313. Naissance ; 409. Vérité ; 478. Philosophie) qui sont attachés aux différentes *classes de concepts*. Chaque *item* - comme *existence* - comporte autant de renvois numérotés qu'il compte d'acceptions réparties dans les différentes *catégories* de l'ouvrage. <sup>123</sup> L'index du Thésaurus Larousse réunit environ 100.000 items. <sup>124</sup>

...

**exilien** 640.18

<sup>117</sup> Figure 4 de S. Reusch, élaboré d'après les informations trouvés dans le Thésaurus Larousse (1992), Corpus : Article 1. Existence, p.1.

<sup>118</sup> Cf. Péchoin (2001), p.91.

<sup>119</sup> Cf. Boulanger (1993), p.133 ; cf. Thésaurus Larousse (1992), Préface, p. v.

<sup>120</sup> Cf. Boulanger (1993), p.133 ; cf. Thésaurus Larousse (1992), Préface, p. v.

<sup>121</sup> Cf. Péchoin (2001), p.96.

<sup>122</sup> Figure 5 de S. Reusch, élaboré d'après les informations trouvées dans le Thésaurus Larousse (1992), Index, p. 820.

<sup>123</sup> Cf. Quemada (1967), p. 386.

<sup>124</sup> Cf. Boulanger (1993), p. 134.

**exine** 289.3  
**existant** 1.3 ; 1.12 ; 409.20 ; 478.19

<b>existence 1</b> vie 310.7 terme de philosophie 409.2 ; 478.19 <i>arriver à l'existence</i> 1.10 <i>recevoir l'existence</i> 313.18
---

**existential** 1.12  
**existentialiser** 1.11  
**existentialisme** 1.2; 478.15; 516.12

...

### Figure 5 : L'index du Thésaurus Larousse pour l'item "existence"

Le thésaurus est un dictionnaire d'analogie qui permet un passage *des idées aux mots* et *des mots aux idées*. Le *Plan de classification* et l'*index alphabétique* sont les « cordons ombilicaux »<sup>125</sup> du thésaurus qui permettent un repérage d'informations dans deux directions. Le *Plan de classification* permet d'explorer à partir d'une *idée* les *mots* qui s'y rattachent. L'*index alphabétique* permet de trouver les *idées* à partir d'un *mot* qui les exprime.<sup>126</sup> Un dictionnaire analogique standard, comme le '*Boissière*' (1862) ou le '*Dictionnaire des analogies*' (2009), fonctionne dans une seule direction. Il ordonne les *mots* selon leurs différentes configurations linguistiques, par ordre alphabétique.<sup>127</sup> Le thésaurus se caractérise par sa bidirectionnalité : Le *Plan de classification* et l'*index alphabétique* font apparaître l'ouvrage comme une somme des *signes* dont les *facettes signifiantes* et *signifiées* sont en interaction constante.<sup>128</sup> Dans cette perspective, une étude de la structure et du fonctionnement d'un thésaurus tend à donner une interprétation précise de la notion vague de « classement sémantique ». <sup>129</sup>

## 2. Le Thésaurus Larousse - La fonction classificatoire des concepts

### 2.1 Définition de l'antonymie

Le *Thésaurus Larousse* fonde sa structure classificatoire sur une relation antonymique entre *termes*. La *hiérarchie des notions* organise les descripteurs par paires opposées, ce qui permet d'associer sens et contre-sens d'une notion conceptuelle (cf. tableau 12). Les

<sup>125</sup> Cité d'après Boulanger (1993), p. 134.

<sup>126</sup> Cf. Thésaurus Larousse (1992), Préface, p. v.

<sup>127</sup> Cf. Boulanger (1993), p. 134.

<sup>128</sup> Cf. Boulanger (1993), p. 134.

<sup>129</sup> Cf. Quemada (1967), pp. 360/361.

antonymes sont des termes <sup>130</sup> (comme 446. *Patience* et 447. *Impatience*) qui appartiennent *obligatoirement* à une même catégorie conceptuelle (comme B.III *Esprit*) et s'opposent par un ou plusieurs traits sémantiques, ayant d'autres en commun.<sup>131</sup>

**Tableau 12 : Le classement par descripteurs opposés dans le Thésaurus Larousse** <sup>132</sup>

A. LE MONDE	B. L'HOMME	C. LA SOCIÉTÉ
I. Grandes notions	III. Esprit	I. Les relations sociales
1. Existence	2. Affectivité	1. Relations
1 Existence	440 Sensibilité	581 Sociabilité
2 Inexistence	441 Insensibilité	582 Insociabilité
3 Matérialité	442 Optimisme	583 Compagnie
4 Immatérialité	443 Pessimisme	584 Solitude
[...]	444 Entrain	585 Bonté
11 Apparition	445 Paresse	586 Méchanceté
12 Disparition	446 Patience	587 Générosité
	447 Impatience	588 Égoïsme

Cette organisation par l'antonymie est intéressante pour la conception logique d'un thésaurus, car elle reflète la tendance générale de l'homme de catégoriser l'expérience et les connaissances du monde en termes de contrastes dichotomiques.<sup>133</sup> La notion de *contraste* implique un alignement symétrique de termes à base d'une ou de plusieurs propriétés partagées, au sein de l'ensemble ordonné dans lequel ils figurent, comme montre l'exemple des termes attachés à la division 2. *Affections* de la catégorie B.III *Esprit* (cf. figure 6)<sup>134</sup>.

<sup>130</sup> Tous de la classe grammaticale de noms.

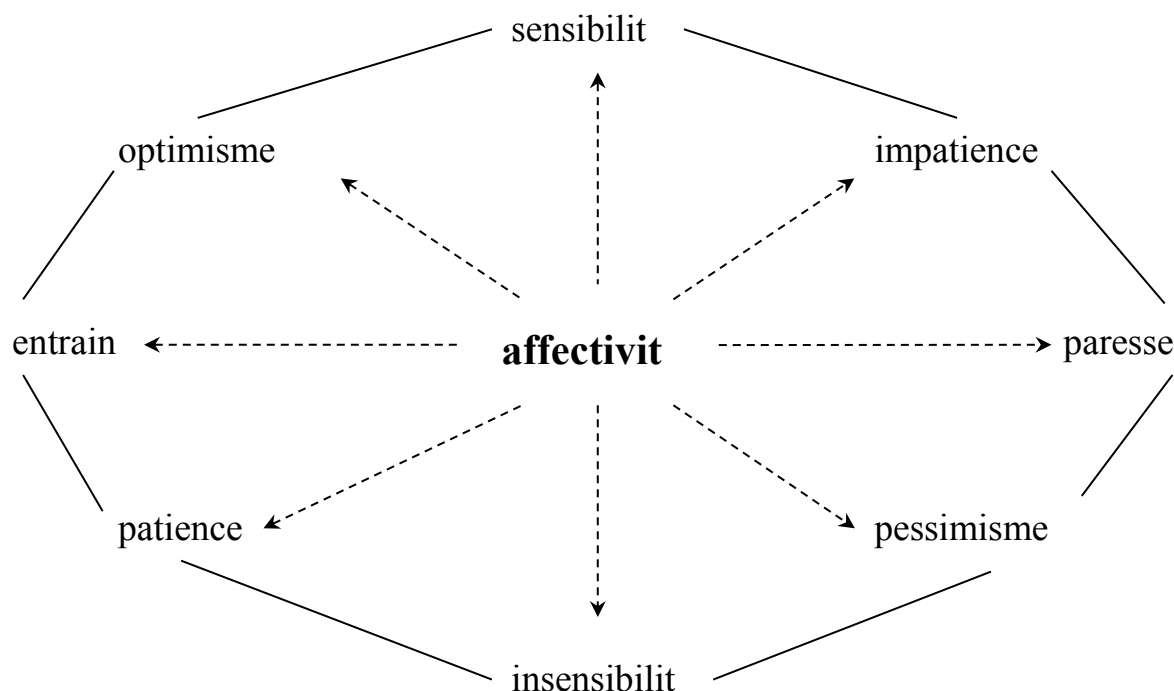
<sup>131</sup> Cf. Dictionnaire Larousse (1991), Article : « Antonymie », p. 40.

<sup>132</sup> Tableau 12 de S. Reusch, élaboré d'après les informations trouvées dans le Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.

<sup>133</sup> Cf. Niklas-Salminen (1997), p. 114.

<sup>134</sup> Cf. Dictionnaire de linguistique (1994), Article : Antinomie, p. 40 ; cf. Schwab et al. (2002), p.2 ; cf. Schwab (2005), p. 100.

Figure de S. Reusch d'après les informations tirés du Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.



**Figure 6 : Alignement symétrique des termes figurant au Trésor Larousse**

L'alignement des termes s'opère de différentes manières, selon la nature du support symétrique.<sup>135</sup> Dans le contexte de thésauri, on étudie un classement qui distingue quatre possibilités de sens contraire<sup>136</sup> :

- *L'antonymie complémentaire* : Elle porte sur l'application et la non-application d'une propriété (exemple : présence/absence)
- *L'antonymie scalaire* : Elle porte sur une propriété de valeur étalonnable (exemple : chaleur/froideur)
- *L'antonymie réciproque & l'antonymie duale* : Elles portent sur une propriété ou un élément considéré comme symétrique par l'usage ou par la convention [exemple : étudiant/professeur (réciproque) ; soleil/lune (dual)]

La considération du support symétrique, constitue le fondement pour une analyse mathématique et logique de l'antonymie.

## 2.2 Analyse des quatre types d'antonymes

L'analyse des quatre types d'antonymes se divise en deux étapes. Dans un premier temps, on analysera l'*antonymie complémentaire* et l'*antonymie scalaire*. Ce sont les types

<sup>135</sup> Cf. Schwab et al. (2002b), p.702 ; cf. Schwab et al (2002a), p. 128.

<sup>136</sup> Cf. Schwab et al. (2002b), p.702 ; cf. Schwab et al (2002a), pp. 128/129.

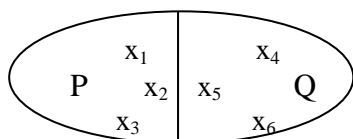
d'antonymes considérés comme « classiques » par la linguistique actuelle. Dans un deuxième temps, on analysera l'*antonymie réciproque* et l'*antonymie duale*. Ces types sont considérés comme « exotiques », voire comme discutables, par certains linguistes. L'étude repose entièrement sur les exemples trouvés dans le *Plan de classification* du Thésaurus Larousse.

### 2.2.1 Antonymie complémentaire

L'antonymie complémentaire concerne, dans le Thésaurus Larousse, les couples comme 'Existence'/'Inexistence' ou 'Présence'/'Absence'.<sup>137</sup> Elle porte sur l'application et la non-application d'une propriété, désignée par les descripteurs :

*Émile est présent.*     $\Rightarrow$  *Émile n'est pas absent.*    *Émile n'est pas présent.*     $\Rightarrow$  *Émile est absent.*  
*Émile est absent.*     $\Rightarrow$  *Émile n'est pas présent.*    *Émile n'est pas absent.*     $\Rightarrow$  *Émile est présent.*

Les deux antonymes s'inscrivent par cette propriété ('existant dans l'espace') dans une relation de *disjonction exclusive*. C'est-à-dire, l'affirmation d'une propriété implique nécessairement la négation de l'autre.<sup>138</sup> L'exemple montre, un objet  $x$  [*Émile*] ne peut pas avoir le prédicat  $P(x)$  [*présent*] et le prédicat  $Q(x)$  [*absent*] à la fois. Ce fait s'explique par le schéma suivant : On a deux ensembles d'objets. Pour l'un vaut le prédicat P, pour l'autre le prédicat Q :<sup>139</sup>



En expressions algébriques<sup>140</sup> :

$$(\forall x) [P(x) \Rightarrow \neg Q(x)] \quad (\forall x) [\neg P(x) \Rightarrow Q(x)]$$

$$(\forall x) [Q(x) \Rightarrow \neg P(x)] \quad (\forall x) [\neg Q(x) \Rightarrow P(x)]$$

<sup>137</sup> Cf. Schwab (2001), p. 13 ; cf. Schwab et al (2002a), p. 128 ; Schwab et al. (2002b), p.703 ; cf. Schwab (2005), p. 101 ; cf. Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.

<sup>138</sup> Cf. Schwab (2001), p. 13 ; cf. Schwab et al (2002a), p. 128 ; Schwab et al. (2002b), p.703 ; cf. Zemmour (2008), p. 165 ; cf. Lehmann /Martin-Berthet (2000), p. 60.

<sup>139</sup> Schéma de S. Reusch

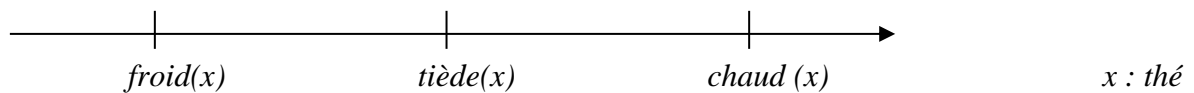
<sup>140</sup> Expression algébrique adoptée d'après Schwab et al. (2002b), p. 703.

Les propriétés (ou prédicats logiques) de descripteurs s'inscrivent dans une relation de symétrie introduite par la négation (indiqué par le signe  $\neg$ ).<sup>141</sup> La négation s'opère, au niveau des termes, par ajout des préfixes grecs ou latins (tel que *ab-*, *dés-*, *dis-*, *in-*, *im-*, *ir-*, *non-*) à un terme de base.<sup>142</sup> Les exemples, qui figurent dans le *Plan de classification* du Thésaurus Larousse, sont les suivants : Présence/ Absence, Espoir/Désespoir, Ressemblance/Dissemblance, Existence/Inexistence, Possibilité/ Impossibilité, Résolution/Irrésolution, Sens/Non-sens, et autres.<sup>143</sup>

### 2.2.2 Antonymie scalaire

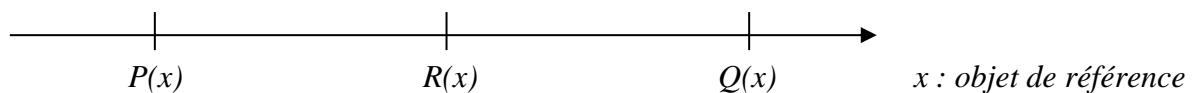
L'antonymie scalaire définit deux descripteurs comme deux « *valeurs limite* » d'une échelle de gradation, qui implique l'existence d'un degré intermédiaire.<sup>144</sup> Elle concerne les systèmes échelonnés comme la taille ('Grandeur'/'Petitesse'), la température ('Chaleur'/'Froid') et l'aridité ('Humidité'/'Sècheresse'), qui figurent dans le *Plan de classification* du Thésaurus Larousse.<sup>145</sup>

*Échelle de la température*<sup>146</sup>



L'antonymie scalaire introduit une symétrie entre deux prédicats logiques [qui sont les propriétés d'objet désignés  $P(x)$  et  $Q(x)$ ], par rapport à une valeur de référence [désigné  $R(x)$ ].<sup>147</sup>

*Échelle de l'antonymie scalaire*<sup>148</sup>



<sup>141</sup> Cf. Schwab (2001), p.13 ; Schwab et al. (2002b), p.703 ; Schwab et al. (2005), p. 74.

<sup>142</sup> Cf. Old (2003), p.17.

<sup>143</sup> Cf. Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.

<sup>144</sup> Cf. Schwab (2001), p. 14 ; cf. Schwab et al. (2005), p.75 ; cf. Lehmann/ Martin-Berthet (2000), p.58.

<sup>145</sup> Cf. Zemmour (2008), p. 165 ; cf. Schwab (2001), p. 14 ; cf. Schwab et al. (2005), p.75 ; cf. Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.

<sup>146</sup> Schéma de S. Reusch

<sup>147</sup> Cf. Schwab (2001), p. p. 14 ; cf. Schwab et al. (2005), p.75.

<sup>148</sup> Schéma de S. Reusch



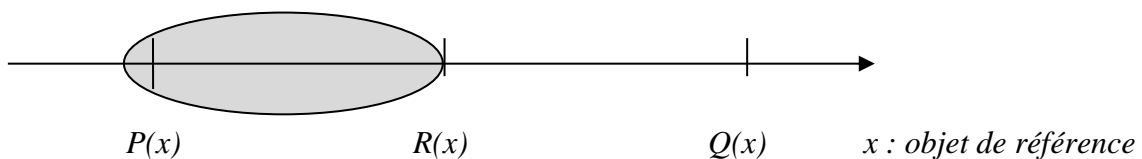
Pour la définition de l'antonymie scalaire nous proposons les expressions algébriques suivantes <sup>149</sup> :

1.  $\forall x P(x) \Rightarrow \neg Q(x)$
2.  $\forall x Q(x) \Rightarrow \neg P(x)$
3.  $\forall x \neg P(x) \Rightarrow Q(x) \vee R(x)$
4.  $\forall x \neg Q(x) \Rightarrow P(x) \vee R(x)$
5.  $\forall x P(x) \Rightarrow \neg Q(x) \vee R(x)$
6.  $\forall x Q(x) \Rightarrow \neg P(x) \vee R(x)$

Si nous reprenons l'exemple de l'échelle de température et considérons que l'élément  $x$  est un thé qui a les propriétés d'être froid :  $P(x)$ , chaud :  $Q(x)$  ou tiède  $R(x)$ , les termes mènent aux expressions algébriques suivantes <sup>150</sup> :

1. *Le thé est froid.*  $\Rightarrow$  *Le thé n'est pas chaud.*
2. *Le thé est chaud.*  $\Rightarrow$  *Le thé n'est pas froid.*
3. *Le thé n'est pas froid.*  $\Rightarrow$  *Le thé est chaud. ou Le thé est tiède.*
4. *Le thé n'est pas chaud.*  $\Rightarrow$  *Le thé est froid. ou Le thé est tiède.*
5. *Le thé est froid.*  $\Rightarrow$  *Le thé n'est pas chaud. ou Le thé est tiède.*
6. *Le thé est chaud.*  $\Rightarrow$  *Le thé n'est pas froid. ou Le thé est tiède.*

La définition logique de l'antonymie scalaire mène à une interprétation spécifique du prédicat pour la valeur intermédiaire  $R(x)$ . Les termes logiques attribuent au prédicat  $R(x)$  un statut particulier en vue des autres prédicats pour la valeur limite,  $P(x)$  et  $Q(x)$ . Le terme  $\forall x P(x) \Rightarrow \neg Q(x) \wedge R(x)$  montre le prédicat comme régissant de l'antonymie scalaire. D'après le terme on attribue à tous les objets du prédicat  $P$  aussi le prédicat  $R$ . Ainsi,  $R(x)$  appartient à la circonférence de  $P(x)$  : <sup>151</sup>

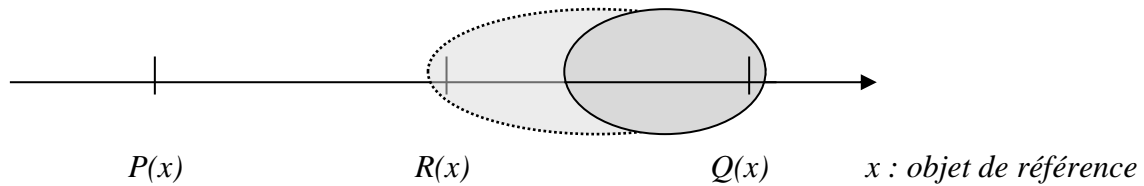


<sup>149</sup> Expressions algébriques de S. Reusch, retravaillé des termes proposés par Schwab et al. (2005), p. 75.

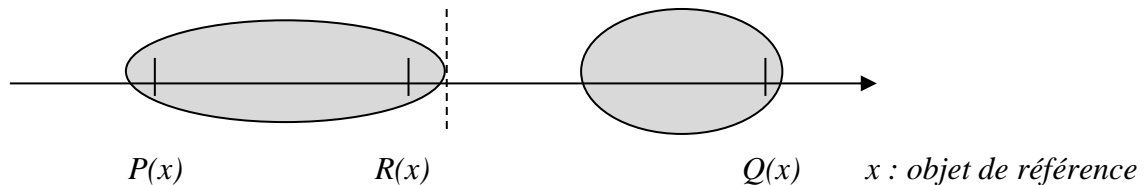
<sup>150</sup> Exemples de S. Reusch, retravaillent des propositions de Schwab et al. (2005), p. 75.

<sup>151</sup> Schéma de S. Reusch

De l'autre côté, on n'attribue pas obligatoirement à chaque objet du prédicat  $Q(x)$  aussi le prédicat  $R(x)$ .  $R(x)$  peut appartenir à la circonférence de  $Q(x)$ , mais pas nécessairement : <sup>152</sup>



Cela veut dire que le type de graduation avec une valeur intermédiaire est dépendant de quel prédicat de base est formé la valeur intermédiaire. Dans ce cas la valeur intermédiaire est plus proche à  $P$  que de  $Q$  : <sup>153</sup>



Dans la définition d'une antonymie scalaire, le rapport avec une échelle de mesure peut aussi être établie. Les échelles de mesure sont généralement utilisées et applicables de manière naturelle, en tant qu'échelle de température avec des valeurs en degrés Celsius, Kelvin, Réaumur ou la mesure classique de la taille en pouces ou en centimètres. Pour la classification des prédicats «chaud» ou «froid » à un objet  $x$ , on peut se référer à des valeurs des objets  $x, y$  associées à une fonction de mesure  $M$ . Les valeurs mesurées se situent dans un intervalle  $I$  de nombres réels : <sup>154</sup>

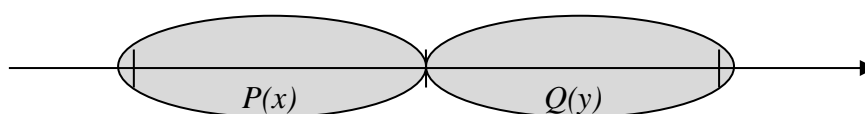
$$M: \{x, y, \dots\} \mapsto I \subset \mathbb{R}$$

Sur cet intervalle, l'on place la limite  $G$ ; les objets ont la propriété  $P(x)$ , si et seulement si,

$$M(x) < G$$

et la propriété  $Q(y)$ , si et seulement si,

$$M(y) > G.$$



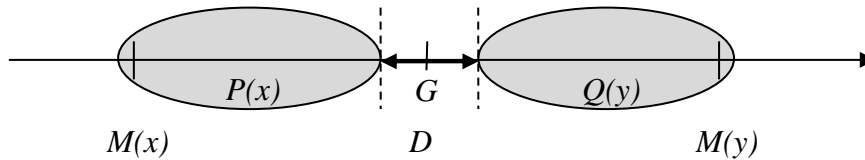
<sup>152</sup> Schéma de S. Reusch

<sup>153</sup> Schéma de S. Reusch

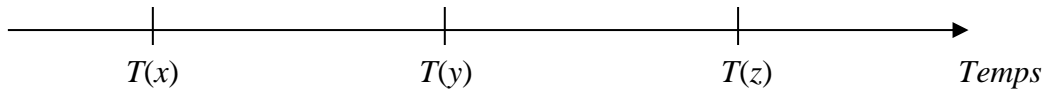
<sup>154</sup> Expressions algébriques et schéma de S. Reusch

$M(x)$  $G$  $M(y)$ 

C'est dans la nature des choses que l'on a besoin pour chaque antonymie scalaire, d'une échelle, d'une fonction de mesure, et d'une valeur limite. Si les antonymies scalaires sont encore plus prononcées, alors, une distance minimale  $D$  entre deux objets  $x$  et  $y$  peut aussi être fixée.<sup>155</sup>



L'échelle de mesure est applicable aussi à la mesure du temps. Une fonction peut placer trois événements  $x$ ,  $y$  et  $z$  sur une axe du temps pour indiquer leur apparition à un moment précis  $T$ .<sup>156</sup>

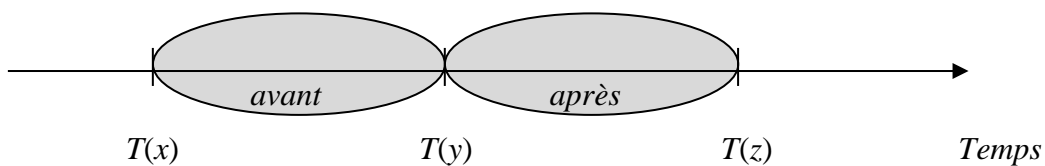


Les trois événements forment un cadre pour une définition de la notion 'avant'/'après'. La fonction suivante montre la définition :<sup>157</sup>

$T: \{x, y, z, \dots\} \mapsto I \subset \mathbb{R}$

$x$  est *avant*  $y$ , si et seulement si :  $T(x) < T(y)$

$z$  est *après*  $y$ , si et seulement si :  $T(z) > T(y)$



La définition de la notion 'avant'/'après' est aussi possible à l'aide de deux prédicats  $T(x)$  et  $T(y)$ , car on peut dire :<sup>158</sup>

Le matin ( $x$ ) est *avant* le soir ( $y$ ).  $\rightarrow$  *avant* ( $x, y$ ),

<sup>155</sup> Schéma de S. Reusch

<sup>156</sup> Schéma de S. Reusch

<sup>157</sup> Expressions algébriques et schéma de S. Reusch

<sup>158</sup> Expressions algébriques et schéma de S. Reusch

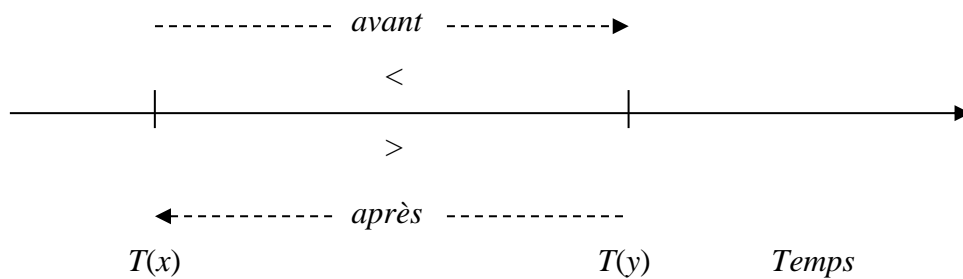
Le soir (y) est *après* le matin (x).  $\rightarrow$  *après* (y,x).

En termes logiques :

$(\forall x,y) [\textit{avant} (x,y) \Leftrightarrow \textit{après} (y,x)]$

$(\forall x,y) : \textit{avant} (x,y) \Leftrightarrow T(x) < T(y)$

$(\forall x,y) : \textit{après} (y,x) \Leftrightarrow T(x) > T(y)$



### 2.2.3 Antonymie réciproque

L'antonymie réciproque <sup>159</sup> concerne les couples comme 'Élève'/'Instituteur', 'Achat'/'Vente' ou 'Entrée'/'Sortie', qui figurent dans le *Plan de classification* du Thésaurus Larousse. <sup>160</sup>

De nombreux linguistes, comme Igor Mel'čuk [Mel'čuk et al. 1995], rangent les exemples ci-nommés dans la rubrique de l'antonymie complémentaire ou scalaire. <sup>161</sup> La recherche moderne en Traitement automatique des langues, qui considère l'antonymie comme un *mécanisme de symétrie* traite la réciprocité comme une antonymie « étendue », de symétrie spécifique. <sup>162</sup> Dans cette perspective, un *élève* est l'opposé symétrique d'un *instituteur* :

Jean est l'*élève* de Pierre.  $\Leftrightarrow$  Pierre est l'*instituteur* de Jean.

En expressions algébriques:

$(\forall x,y) [P(x,y) \Leftrightarrow Q(y,x)]$ . <sup>163</sup>

<sup>159</sup> Egalement appelé « antonymie converse », cf. Schwab (2001), p. 14/15 ; cf. Schwab et al. (2005), pp. 75/76 ; cf. Lehmann /Martin-Berthet (2000), pp. 60/61.

<sup>160</sup> Cf. Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.

<sup>161</sup> Cf. Schwab et al. (2005), p. 75 ; cf. Schwab (2001), p. 14/15.

<sup>162</sup> Cf. Schwab et al. (2005), pp. 75/76.

<sup>163</sup> Expression algébrique citée d'après Schwab et al. (2005), p.75.

La symétrie spécifique se montre quand on lit l'expression logique. Pour l'exemple présent la formule dit : Pour tous les personnes (x,y), x est élève de y, si et seulement si, y est instituteur de x. La symétrie est fondée sur une forte dépendance de deux termes qui sont liés dans le cas présent par une relation hiérarchique spécifique de l'institution scolaire.

Selon nos considérations, il existe deux types d'antonymes réciproques au sein du Thésaurus Larousse. Un premier type recourt à un objet de référence qui est intuitivement donné. Parmi ce type comptent des couples tels que 'étudiant'/'instituteur' ou 'maître'/'serviteur'. Le second type nécessite la précision d'un objet de référence concret. Parmi ce type comptent les couples désignant un échange des biens tels que 'achat'/'vente', 'prêt'/'emprunt', 'don'/'réception' et les couples désignant une relation spatiale, tels que 'devant'/'derrière' et 'entrée'/'sortie'.<sup>164</sup> Dans les deux cas l'objet de référence (le bien échangé ou la référence spatiale définie) doit être précisé. Les exemples suivants expliquent les deux types d'antonymie réciproque à l'aide d'une expression algébrique.

Exemple 1 : 'achat'/'vente'<sup>165</sup>

$(\forall x,y,z) [acheter(x,y,z) \Leftrightarrow vendre(y,x,z)]$

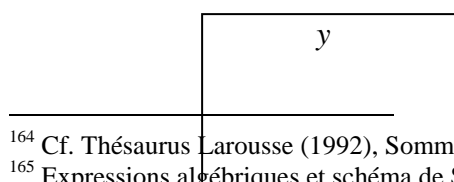
Jean (x) achète à Pierre (y) *un livre* (z).

Pierre (y) vend à Jean (x) *un livre* (z).

La relation antonymique 'achat'-'vente' se définit comme un échange de biens entre deux personnes (x,y). L'expression logique nécessite un prédicat supplémentaire (z), qui précise l'objet de référence auquel la relation peut être rapportée.

Exemple 2 : 'entrée'/'sortie'<sup>166</sup>

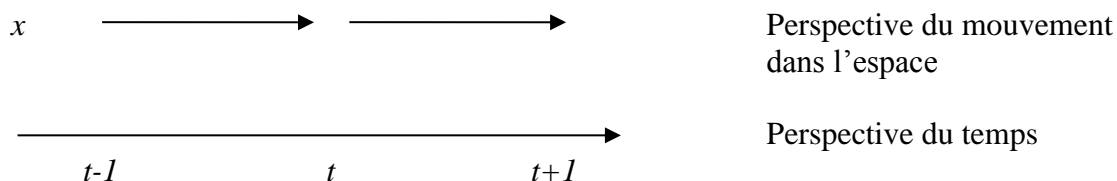
La relation antonymique 'entrée'-'sortie' présente une symétrie qui est basée sur une perspective spatio-temporelle sur l'action humaine. Elle nécessite plusieurs objets de référence : Une personne (x), un espace fermé (y) et une mesure du temps (t) :



<sup>164</sup> Cf. Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.

<sup>165</sup> Expressions algébriques et schéma de S. Reusch

<sup>166</sup> Expressions algébriques et schéma de S. Reusch.



En termes logiques :

$(\forall x) : \exists y \quad \text{entrer}(x,y) \Leftrightarrow x \in y \text{ à un moment } t \wedge x \notin y \text{ à un moment } t-1$

$(\forall x) : \exists y \quad \text{sortir}(x,y) \Leftrightarrow x \in y \text{ à un moment } t \wedge x \notin y \text{ à un moment } t+1$

1) Pour tous les  $x$  vaut : Il existe un *espace fermé* ( $y$ ). On ne considère que  $x$  est élément d'un espace fermé à un moment  $t$ . Et, que  $x$  n'est pas élément d'un espace fermé à un moment  $t-1$ . *Entrer* signifie dans cette logique : Une personne ( $x$ ) entre dans l'espace fermé à un moment  $t$ , si elle n'est pas dans l'espace à un moment  $t-1$ , mais qu'elle se trouve dans l'espace fermé à un moment  $t$ .

2) Pour tous les  $x$  vaut : Il existe un *espace fermé* ( $y$ ). On ne considère que  $x$  est élément d'un espace fermé à un moment  $t$ . Et, que  $x$  n'est pas élément d'un espace fermé à un moment  $t+1$ . *Sortir* signifie dans cette logique : Une personne  $x$  sort d'un espace fermée à un moment  $t$ , si elle est dans l'espace fermé à un moment  $t$  et ne se trouve plus dans l'espace fermé à un moment  $t+1$ .

#### 2.2.4 Antonymie duale

L'antonymie duale met en relation deux descripteurs par une symétrie qui porte sur des *fonctions culturelles* (symétrie consacrée par l'usage) et sur des *fonctions spatio-temporelles* (symétrie consacrée par les propriétés spécifiques de l'espace-temps).<sup>167</sup>

L'antonymie duale modélise des relations diverses. Le classement sémantique du Thésaurus Larousse présente trois types d'antonymes duals:

- Les mots que la culture associe (ex. '*Soleil*'/'*Lune*'),
- Les mots qui dépendent logiquement l'un de l'autre (ex. '*Question*'/'*Réponse*'),
- Les mots qui expriment le passage d'un état à un autre (ex. '*Naissance*'/'*Mort*').<sup>168</sup>

Dans le cas présent, nous proposons une étude du couple '*Naissance*'/'*Mort*'. Pour sa description logique nous proposons l'expression algébrique suivante :

<sup>167</sup> Cf. Schwab (2001), p. 15 ; cf. Schwab et al. (2002a), pp.128/129 ; cf. Schwab et al. (2005), p. 76.

<sup>168</sup> Cf. Schwab (2001), p.15 ; cf. Schwab et al. (2002a), p. 128 ; cf. Schwab et al. (2005), p. 76. Exemples tirés du Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.

$$\exists x \exists t_1 [P(x, t_1)] \Leftrightarrow \exists x \exists t_2 [Q(x, t_2)] \wedge t_1 < t_2. \quad ^{169}$$

Cette formule signifie : Il existe un élément  $x$  et un moment  $t_1$ . L'élément  $x$  peut désigner tout être animé, auquel on peut appliquer le prédicat vivant, alors que le moment  $t_1$  peut désigner l'état initial de  $x$ , donc la naissance. Dans cette logique, un être vivant ( $x$ ) meurt à un moment  $t_2$ , si et seulement si, il est né à un moment  $t_1$ , un moment qui se situe antérieurement au moment  $t_2$  (décrit par  $t_1 < t_2$ ).

Nous considérons que l'antonymie duale ne modélise pas uniquement le rapport de deux valeurs par rapport à un objet  $x$ . Elle modélise aussi la relation de symétrie qui s'instaure entre deux objets  $x$  et  $y$ , qui ont une valeur propre  $P(x)$  et  $Q(y)$ . Nous adoptons ici l'expression logique proposée par Schwab et al. (2005) :

$$(\exists x) [P(x)] \Leftrightarrow (\exists Q \exists y) [Q(y)]. \quad ^{170}$$

Cette expression algébrique exprime le rapport entre les mots qui dépendent logiquement l'un de l'autre, tels que '*Question*'/'*Réponse*'. Elle exprime la relation antonymique entre '*Question*' et '*Réponse*' de manière suivante : Si  $x$  est une question [ayant le prédicat  $P(x)$ ], il existe aussi un objet  $y$  et un prédicat  $Q$  lié à cet objet [désigné par  $Q(y)$ ] qui est la réponse. Par conséquent,  $Q(y)$  est la *réponse* à la *question*  $P(x)$ .<sup>171</sup> Les exemples choisis montrent la forte dépendance logique de deux prédicats qui est typique pour l'antonymie duale, voire l'antonymie réciproque.

L'analyse mathématique et logique montre l'importance de la notion de « symétrie » pour une définition précise de ce que sont les *deux facettes* de l'antonymie au sein du Trésor Larousse :

Sur le *plan structurel*, on peut parler d'une antonymie, si :

- deux *termes génériques* appartiennent à un même *domaine de sens* (indiqué par l'appartenance des termes à une même *catégorie*) et se placent sur un même *axe* lors d'une comparaison (indiqué par une relation de *co-hyponymie* entre les termes).<sup>172</sup>

Sur le *plan sémantique*, on peut parler d'une antonymie, si :

<sup>169</sup> Expression algébrique de S. Reusch.

<sup>170</sup> Expression algébrique adaptée d'après Schwab et al. (2005), p. 76.

<sup>171</sup> Cf. Schwab (2001), p. 15 ; cf. Schwab et al. (2005), p. 76.

<sup>172</sup> Cf. Schwab (2005), p. 100 ; cf. Sémanticlopédie, Article « antonymie » (de Pierre Amisili), consultable sur internet.

- deux *termes spécifiques* ont certains *sèmes* en commun et si leur *sèmes* non communs sont en opposition.<sup>173</sup>

Cette définition fait apparaître deux facettes de l'antonymie qui sont en interaction constante au sein d'un thésaurus.<sup>174</sup>

## 2.3 Un classement basé sur des notions opposées

Au sein du Thésaurus Larousse, la présence des descripteurs antonymes conditionne étroitement la composition des articles. Les articles sont disposés, partout où cela est possible, par paires de notions opposées ou corrélatives. Selon les rédacteurs : « Cette disposition permet d'embrasser d'un seul coup d'œil l'ensemble des synonymes, des contraires et des analogies d'un mot donné. »<sup>175</sup> Prenons pour exemple les articles 28. *Conformité* et 29. *Non-conformité* du Thésaurus Larousse.<sup>176</sup>

---

<sup>173</sup> Cf. Lehmann / Martin-Berthet (2000), p. 59.

<sup>174</sup> Cf. Sémanticlopédie, Article « antonymie » (de Pierre Amisili), consultable sur internet.

<sup>175</sup> Cité d'après Thésaurus Larousse (1992), Préface, p. vi.

<sup>176</sup> Tableau 13 de S.Reusch, élaboré d'après les informations trouvés dans le Thésaurus Larousse (1992), Corpus : articles 28. Conformité et 28. Non-conformité, pp. 14/15.



### Tableau 13 : Étude de la composition antonymique des articles

28. Conformité		29. Non-conformité	
N 1	<b>Conformité</b> , adéquation, <i>convenance</i> , congruence [litt.], pertinence. - <i>Comparabilité</i> , concordance, correspondance. - <i>cohérence</i> , <i>homogénéité</i> . Identité 15, ressemblance 21, similitude ; uniformité. [...]	N 1	<b>Non-conformité</b> , <i>disconvenance</i> , inadéquation, incongruité ; <i>incompatibilité</i> . - <i>Hétérogénéité</i> , <i>incohérence</i> ; irrégularité. [...]
V 2		2	Différence 23 ; disparité, [...]
V 9	<b>Conformer</b> ; <i>accorder</i> , adapter, ajuster. - <i>cohérer</i> [rare] harmoniser, homogénéiser ; <i>assimiler</i> 15, unifier [...]	V 8	<b>Contrevenir</b> , déroger ; <i>se distinguer</i> , se faire remarquer. - Différer de 23 ; faire exception. - <i>Se reTransbeller</i> 651 ; entrer en dissidence.
10		9	<i>Contraster</i> , discorder [litt.], diverger, s'opposer 18. - Dysfonctionner.
11	Se conformer à, se plier à, se soumettre à 628. - Normaliser, régulariser [...]	10	<i>Désaccorder</i> ; désadapter, désajuster. - Altérer 16, dénaturer. [...]
Adj 12	<i>Concorder</i> , correspondre ; cadrer, coller [fam.], convenir [...]	Adj 11	<b>Non-conforme</b> , <i>disconvenant</i> , inadéquat, inapproprié, incongru ; incompatible. - Hétéroclite, hétérogène ; incohérent. [...]
	<b>Conforme</b> , adéquat, approprié, congru, idoine [sout.],. - Adapté, ajusté ; assorti. - <i>Compatible</i> , [...]		
	<i>Cohérent</i> , équilibré, harmonieux, homogène.		

Dans l'ordre du classement, les articles se suivent. Ainsi, on trouve sur la même page la série des 'synonymes' du nom *conformité* (*convenance, comparabilité, cohérence, homogénéité*) la série des noms de signification opposée (*disconvenance, incompatibilité, incohérence, hétérogénéité*). On trouve la série des analogues du verbe *conformer* (*accorder, assimiler, se soumettre à, concorder*) en face de la série analogique du verbe *contrevenir* (*désaccorder, se distinguer, se rebeller, contraster*). On aperçoit la série des synonymes de l'adverbe *conforme* (*approprié, adéquat, congru, compatible, cohérent, homogène*) en face de la série attaché à l'adverbe *non-conforme* (*inapproprié, inadéquat, incongru, incompatible, hétérogène*). D'autres exemples peuvent être trouvés.

La conception du Thésaurus Larousse repose sur l'hypothèse selon laquelle l'appel des traits sémantiques du *terme générique* génère l'énumération d'une série des *termes spécifiques* corrélés dans la langue. Le *terme générique* fonctionne comme un *noyau*

autour duquel on rassemble les unités lexicales pour les déterminer en vue d'une *notion conceptuelle*. Le classement des *termes génériques* par notions opposées permet d'instaurer une *symétrie* entre les traits sémantiques des *termes spécifiques* qui figurent dans les articles connexes. Cette *symétrie* détermine la structure des ces articles. Elle rapproche les *termes spécifiques* en opposition, même s'ils s'inscrivent simultanément dans d'autres relations sémantiques.

### 3. Le Thésaurus Larousse - L'aspect relationnel des concepts

Le corpus du thésaurus se structure selon un réseau de relations sémantiques, qui comprend trois rapports qui expriment la parenté sémantique : La similarité, l'analogie et la synonymie.

Lors de l'utilisation d'un thésaurus, ces relations sémantiques permettent en premier lieu de situer le *descripteur* (ou *terme générique*) dans son environnement conceptuel.<sup>177</sup> L'utilisateur peut ainsi discerner plus facilement le *centre d'idées* et découvrir ce que le système des renvois explicites et implicites peut lui offrir en informations supplémentaires. La structure sémantique lui offre la possibilité de consulter un thésaurus de façon logique à partir de n'importe quel *terme*.<sup>178</sup> Pour cette raison, les '*descripteurs orphelins*'<sup>179</sup> – c'est-à-dire des descripteurs qui ne sont rattachés à aucun autre descripteur – sont considérés comme des intrus au système *thésaurus*. Isolés, ces termes ne mènent à aucune idée nouvelle et ne peuvent être retrouvés à partir d'aucun autre descripteur.<sup>180</sup> L'importance du lien sémantique pour la description d'un terme fait comprendre la *densité du réseau des relations sémantiques* au sein d'un thésaurus. En moyenne, il y a dans un thésaurus près de trois relations sémantiques par *descripteur* (ou *terme générique*).<sup>181</sup> Un thésaurus se présente ainsi comme un ensemble d'unités sémantiques qui représentent chacune – vis-à-vis du *descripteur* - un contexte particulier d'occurrence des termes.<sup>182</sup>

La définition claire des liens sémantiques constitue un enjeu majeur dans la réalisation d'un futur thésaurus. Dans le chapitre présent nous nous intéresserons donc aux différentes approches de la définition et de l'analyse des relations sémantiques. La *similarité sémantique* et l'*analogie* sont à base d'une compréhension approfondie du classement

<sup>177</sup> Cf. Documents numériques de l'UQAM.

<sup>178</sup> Cf. Document "Politiques d'indexation et de gestion de vocabulaire"

<sup>179</sup> Termes cités d'après "Politiques d'indexation et de gestion de vocabulaire"

<sup>180</sup> Cf. Document "Politiques d'indexation et de gestion de vocabulaire"

<sup>181</sup> Cf. Document "Politiques d'indexation et de gestion de vocabulaire"

<sup>182</sup> Cf. Zargayouna / Salotti (2004), p.2.

sémantique opéré dans un thésaurus. Leur différence repose sur le classement de deux termes en catégories identiques, connexes ou à part. Le chapitre présent commence donc par une définition démarcative de ces notions importantes. Cette définition va être approfondie, par la suite, par une '*analyse formelle des concepts*'. La *similarité sémantique* décide en grande partie de la classification de termes. L'entrée de deux termes à une distance proche ou éloignée se décide en fonction d'une parenté sémantique apparente ou non. Par conséquence logique, la position de deux termes dans le corpus peut donner des renseignements sur leur *similarité sémantique*. La structure spécifique d'un thésaurus - qui ressemble à un graphe conceptuel en mathématique - permet d'opérer plusieurs mesures pour évaluer la similarité sémantique entre termes. Ce chapitre présente plusieurs mesures, explique leur fondement logique en explicitant la *théorie des graphes* et évalue les mesures en fonction de leur pertinence pour le Thésaurus Larousse. Après avoir explicité les grandes relations sémantiques que contient un thésaurus, les dernières pages du chapitre sont consacrées à l'analyse de la *synonymie*, une notion centrale en lexicographie.

### 3.1 Analyse formelle du classement par concepts

#### 3.1.1 Définition du concept : Par l'*extension* et l'*intension*

Un thésaurus - comme le Thésaurus Larousse - est un dictionnaire qui est structuré par des concepts, plutôt que par des mots. L'analyse des concepts est donc nécessaire dans le contexte d'une compréhension de l'organisation *explicite* et *implicite* de l'ouvrage.

Un *concept* peut être compris philosophiquement comme l'*unité de base de la pensée* formé dans un processus dynamique au sein des environnements sociaux et culturels. Selon la tradition philosophique, un concept est constitué par son *extension*, comprenant tous les objets qui appartiennent à ce concept, et par son *intension*, comprenant tous les attributs (propriétés et significations) qui s'appliquent à tous les objets de l'extension.<sup>183</sup> Ainsi, un concept existe uniquement dans une relation avec nombreux autres concepts. Dans ce cadre la relation '*subconcept* - *superconcept*'<sup>184</sup> joue (comme la relation

<sup>183</sup> Cf. Depecker (1999), pp. 42/43 ; cf. Wille (2005), p.2 ; cf. Taddei Elmi (1981), pp. 847/848 ; cf. Rey (1976), pp. 84-86.

<sup>184</sup> Cf. Wille (2005), p.2.

Terminologie anglaise adopté de la publication « Formal Concept Analysis as Mathematical Theory » (2005) de Rudolf Wille, qui traduit la relation entre un concept générique et un concept spécifique. La terminologie a été choisie pour ne pas créer des confusions avec les « termes génériques » et « termes spécifiques », présentés dans chapitre 1.3.2.

d'hyponymie/hyperonymie dans le contexte de sémantique lexicale) un rôle important, en ce sens que : Être un '*subconcept*' d'un '*superconcept*' signifie que l'*extension* du '*subconcept*' est contenue dans celle du '*superconcept*'. Des objets comme «bibliothèque » ou « musée » qui sont contenus dans le '*subconcept*' « bâtiment public » sont donc également contenus dans l'*extension* du '*superconcept*' « bâtiment ». Cela implique que l'*intension* du '*subconcept*' contient l'*intension* du '*superconcept*'.<sup>185</sup> Des attributs comme /stationnaire/ ou /bâti par l'homme/ qui sont inhérent à «bâtiment », se trouvent également dans les propriétés qui caractérisent un « bâtiment publique », tel qu'un « musée ». <sup>186</sup>

### 3.1.2 Analyse formelle des concepts

L'*analyse formelle des concepts* (AFC) [Wille 1982] utilise des méthodes mathématiques pour élaborer et analyser les relations entre concepts et les implications hiérarchiques sur des données fournies. <sup>187</sup> L'AFC permet de parler des *objets*, des *attributs* et des *relations* (ou *incidences*) qui indiquent qu'un objet dispose d'un attribut. Et, elle permet d'étudier comment des *objets* peuvent être regroupés hiérarchiquement en fonction de leurs *attributs* communs. La méthode se base donc sur la mathématisation de la compréhension philosophique de ce qui est un *concept* (cf. chapitre 3.1.1). <sup>188</sup> L'*analyse formelle des concepts* utilise pour la formalisation d'un concept, soit une théorie algébrique des relations binaires, soit des treillis complets. <sup>189</sup> La *théorie des relations binaires* présente un *concept* dans un tableau de valeurs, qui définit un *contexte formel*  $K(G,M,I)$ . <sup>190</sup> Le contexte formel est composé d'un ensemble d'*objets* (G) et d'un ensemble d'*attributs* (M) dont il exprime une relation binaire, l'*incidence objet-attribut* (I). <sup>191</sup> La visualisation d'un contexte formel se présente comme un tableau croisé, qui indique par une croix la relation d'incidence attribut-objet (cf. tableau 14) <sup>192</sup> ...

<sup>185</sup> Cf. Wagner (1973), p. 201 ; cf. Wille (2005), p. 2.

<sup>186</sup> Exemples de S.Reusch, tirés du *Thésaurus Larousse* (1992), article 777. Architecture, pp. 535/536.

<sup>187</sup> Cf. Wille (1987), Wille (2005), Koester (2006). Dans les sources, la théorie est communément appelé *Formal concept analysis* (FCA), cf. Koester (2006), p. 13.

<sup>188</sup> Cf. Koester (2006), p. 13 ; Wille (2005), p. 2.

<sup>189</sup> Cf. Koester (2006), pp. 13-21.

<sup>190</sup> La notation du contexte formel  $K(G,M,I)$  est dérivée de la terminologie officielle en allemand : Kontext (K), Gegenstand (G), Merkmal (M), Inzidenz (I), qui pourrait être traduit en français par  $C(O,A,I)$  : Contexte (C), Objet (O), Attribut (A), Incidence (I).

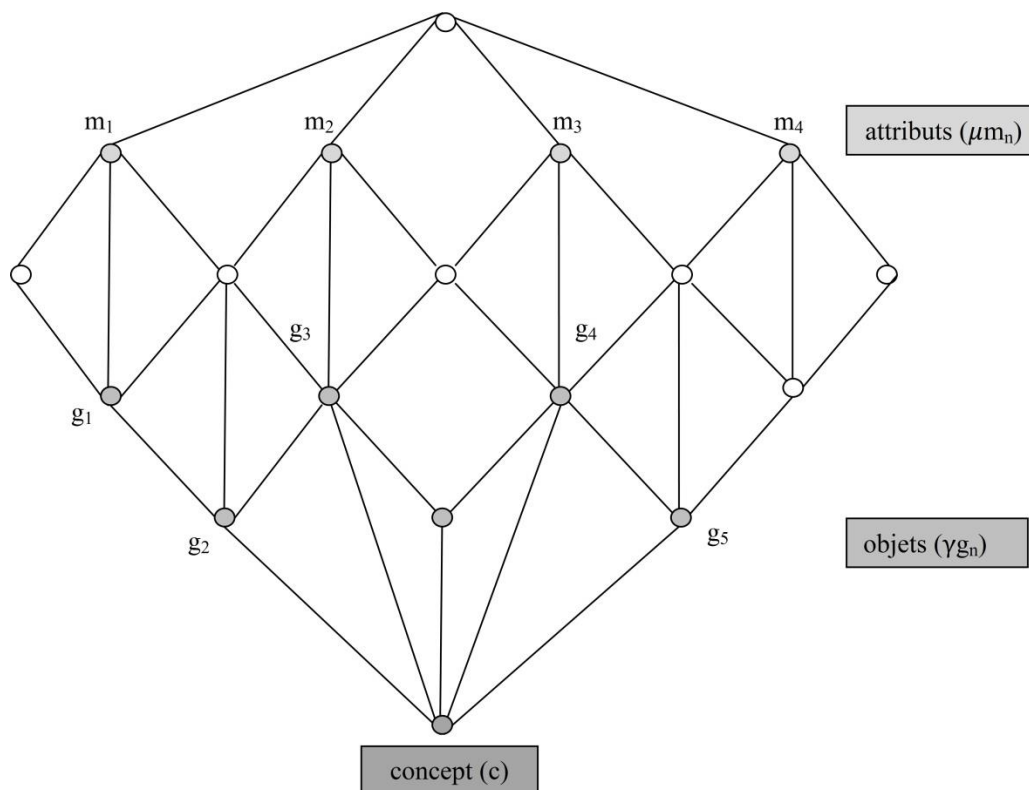
<sup>191</sup> Cf. Koester (2006), p. 14.

<sup>192</sup> Cf. Koester (2006), p. 14 ; Caspard et al. (2007), pp. 3/4.

**Tableau 14: Tableau exemplaire pour l'analyse formelle des concepts** <sup>193</sup>

attributs ( $m_n$ )				
objets ( $g_n$ )	$m_1$	$m_2$	$m_3$	$m_4$
$g_1$	x			
$g_2$		x		
$g_3$			x	
$g_4$				x
$g_5$				x

... ou un *treillis*, qui indique par une connexion la relation d'incidence attribut-objet (cf. figure 7) <sup>194</sup>.



**Figure 7: Treillis exemplaire pour l'analyse formelle des concepts** <sup>195</sup>

<sup>193</sup> Tableau 14 de S. Reusch, élaboré d'après les informations trouvés dans Koester (2006), p. 14 et Caspard et al. (2007), pp. 3/4.

<sup>194</sup> Cf. Caspard et al. (2007), pp. 5-11.

<sup>195</sup> Figure 7 de S. Reusch, élaboré d'après les informations trouvés sur le site officiel de la méthode « Formal Concept Analysis », <http://www.upriss.org.uk/fca/fca.html>, crée par Uta Priss.

Dans un tel treillis le terme de chaque objet  $g_n$  est attaché à son objet représenté  $\gamma g_n$  et le terme de chaque attribut  $m_n$  est attaché à son attribut représenté  $\mu m_n$ .<sup>196</sup> D'après le théorème de base, cet étiquetage permet d'étudier l'*extension*, l'*intension* et le *contexte formel* sous-jacent du concept. Le concept est représenté par un cercle de sorte que son *extension* (*intension*) est composé de tous les *objets* (*attributs*) dont les termes peuvent être atteint par un *chemin descendant* (*ascendant*) de ce cercle.<sup>197</sup>

L'*analyse formelle des concepts* dans un thésaurus est influencée par le classement sémantique des concepts (cf. figure 8).<sup>198</sup> Lors de la recherche d'un concept, l'utilisateur commence par consulter la hiérarchie des notions - résumée sur les premières pages du dictionnaire. Puis, il recherche dans le corpus l'endroit où est classé le concept. L'utilisateur étudie l'entrée dictionnaire attaché au concept. Celle-ci l'informe sur l'*extension* (objets) et l'*intension* (attributs) du concept. Elle structure les informations dans un tableau de valeurs qui lui montre les *incidences attribut-objet*. Ces *incidences* lui fournissent une grille sur laquelle peut être développé un *treillis*.<sup>199</sup> La création d'un *ensemble ordonné* permet l'analyse de l'*organisation implicite* du thésaurus - au niveau conceptuel - , qui vient compléter l'analyse de l'*organisation explicite*, celle de la *hiérarchie des notions*.<sup>200</sup>

---

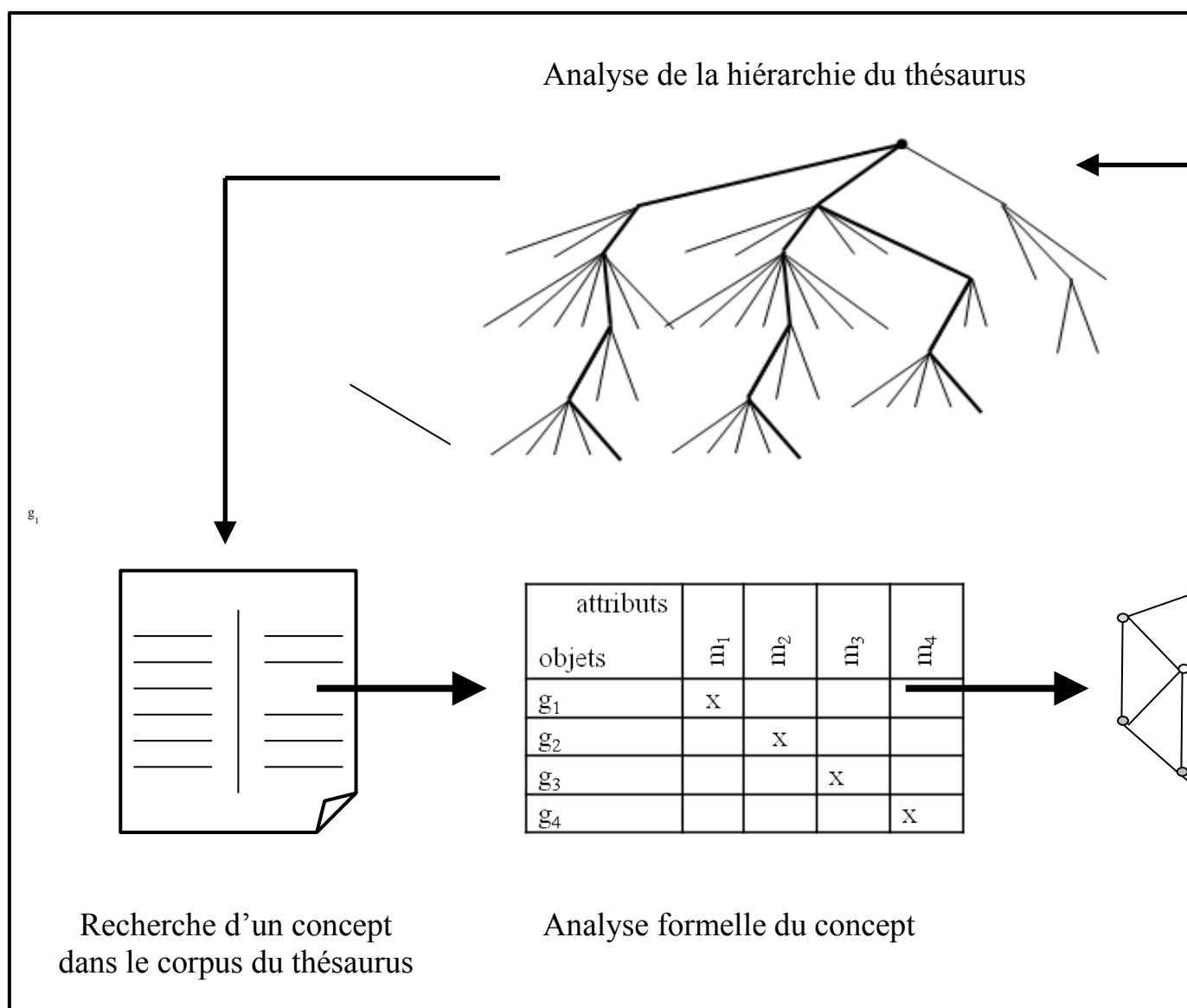
<sup>196</sup> Cf. Wille (2005), p. 3.

<sup>197</sup> Cf. Wille (2005), pp. 3/4.

<sup>198</sup> Cf. Old (2004), pp. 237/238 ; cf. Sedelow (1993), pp. 317-321.

<sup>199</sup> Cf. Old (2004), p. 241.

<sup>200</sup> Cf. Old (2004), p. 237.



**Figure 8: L'analyse formelle d'un concept à l'aide des incidences attribut-objet**<sup>201</sup>

Proposons un exemple. La hiérarchie du Thésaurus Larousse présente une catégorie appelée '*nombre*'. Les termes figurant dans cette catégorie, dénotent les dix premiers nombres naturels, *un*, *deux*, *trois* etc., de notre système de comptage (cf. tableau 10). L'*organisation explicite* présente les termes selon la linéarité. L'analyse formelle permet d'étudier les propriétés des divers nombres. On évalue, si leur nature est */pair/*, */impair/*, */composé/*, */quadratique/*<sup>202</sup> ou si leur nature est celle d'un nombre */premier/*.<sup>203</sup> Le tableau 15 résume les résultats de l'analyse.

<sup>201</sup> Figure 8 de S. Reusch. Le schéma de consultation du Thésaurus Larousse a été construit d'après les expériences faites lors de la création du treillis exemplaire (cf. figure 7) et lors de la création du treillis « nombre » (cf. figure 9).

<sup>202</sup> Quadratique : Terme peu usité, synonyme de « carré ».

---

<sup>203</sup> L'analyse formelle repose sur une sélection des propriétés des nombres entiers . D'autres propriétés, que l'on peut trouver dans le Thésaurus Larousse, peuvent faire l'objet d'une analyse plus détaillée (cf. p.59, article 100. Nombre).

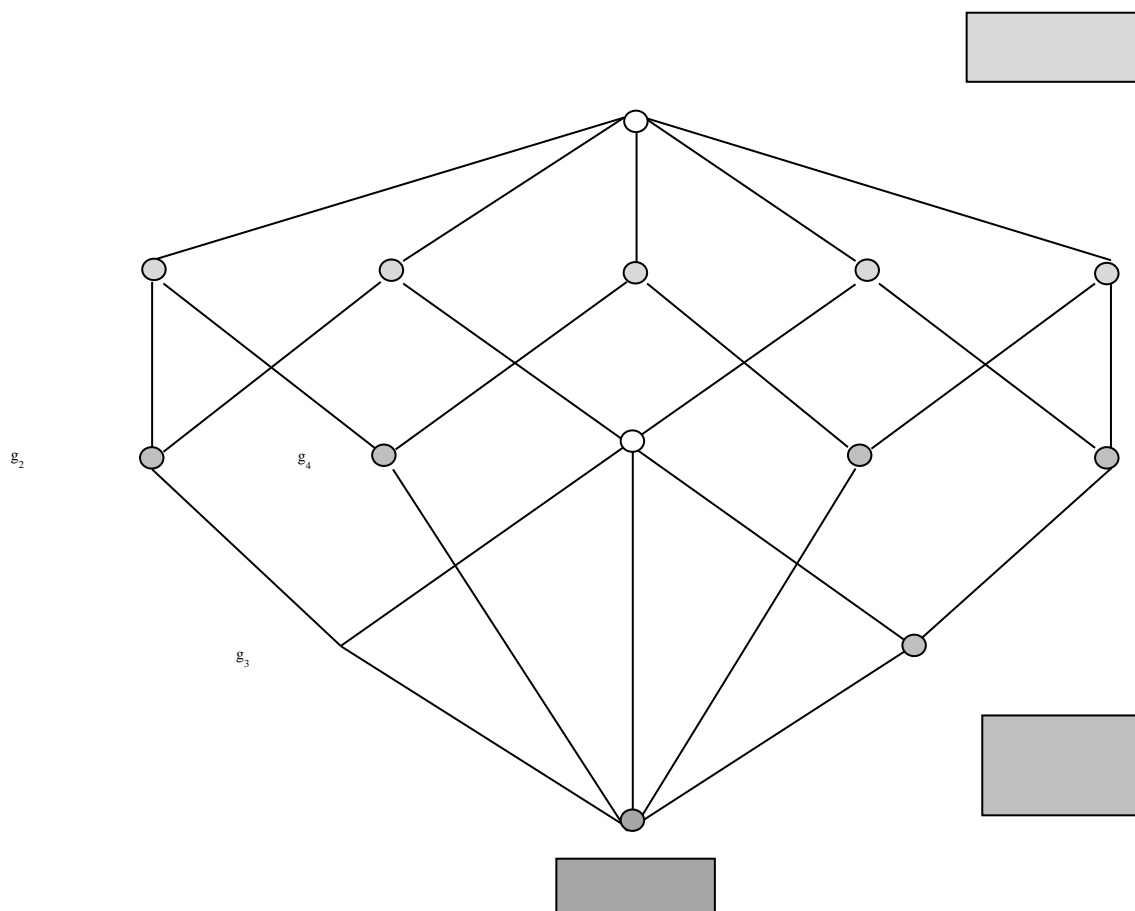


**Tableau 15: L'analyse formelle du concept 'nombre' : La table**<sup>204</sup>

propriété nombre	pair	Impair	premier	composé	quadratique
un		x			x
deux	x		x		
trois		x	x		
quatre	x			x	x
cinq		x	x		
six	x			x	
sept		x	x		
huit	x			x	
neuf		x		x	x
dix	x			x	

Les *propriétés d'objets* créent une structure différente de celle de la *hiérarchie des notions* (cf. figure 9). Regroupées en treillis, elles peuvent former une grille dans laquelle les *objets* sont regroupés ensembles suivant leur *incidence objet-attribut*. Différents nombres, qui partagent une même propriété, forment ainsi un *cercle* (plus précisément un *nœud*). Par exemple, 'trois', 'cinq' et 'sept' représentent le nœud des '*nombres pairs*'. L'*ensemble ordonné* ainsi structuré, montre l'*organisation implicite* du champs conceptuel 'nombre', qui n'est plus basée sur une représentation linéaire, mais sur une représentation logique.

<sup>204</sup> Tableau 15 de S.Reusch, élaboré d'après les informations tirés du Sommaire (pp. xv/xvi, articles 101. Zéro à 111. Dix) et du Corpus (p. 59, article 100. Nombre) du Thésaurus Larousse (1992).



**Figure 9 : L'analyse formelle du concept 'nombre' : Le treillis**<sup>205</sup>

Le Thésaurus Larousse a été formalisé avec des méthodes d'*analyse formelle des concepts*. Il contient un système classificatoire à six niveaux (cf. chapitre 1.3.4). Au niveau le plus élevé de la *hiérarchie des notions* sont organisés les *concepts* (ou *termes génériques*). Au niveau le plus bas sont groupés des *mots* (ou *termes spécifiques*), qui sont soit des synonymes (ou quasi-synonymes), soit des termes étroitement liés selon certain champs sémantiques - comme des termes figurant dans un *article encyclopédique*, qui contient, par exemple, des listes de plantes, d'animaux ou de denrées alimentaires.<sup>206</sup> En raison de leur *polysémie*, beaucoup de mots apparaissent plusieurs fois dans le thésaurus. Ainsi, on peut construire un treillis de concepts en utilisant les *mots* et les *classes de concepts* dans lesquels ils figurent pour étudier les différentes notions d'un concept. Ainsi, on peut construire un treillis de concepts en utilisation les *termes génériques* dans le thésaurus comme *objets* et leurs *termes spécifiques* attachés comme des *attributs* pour étudier les différentes *notions d'un concept*.<sup>207</sup>

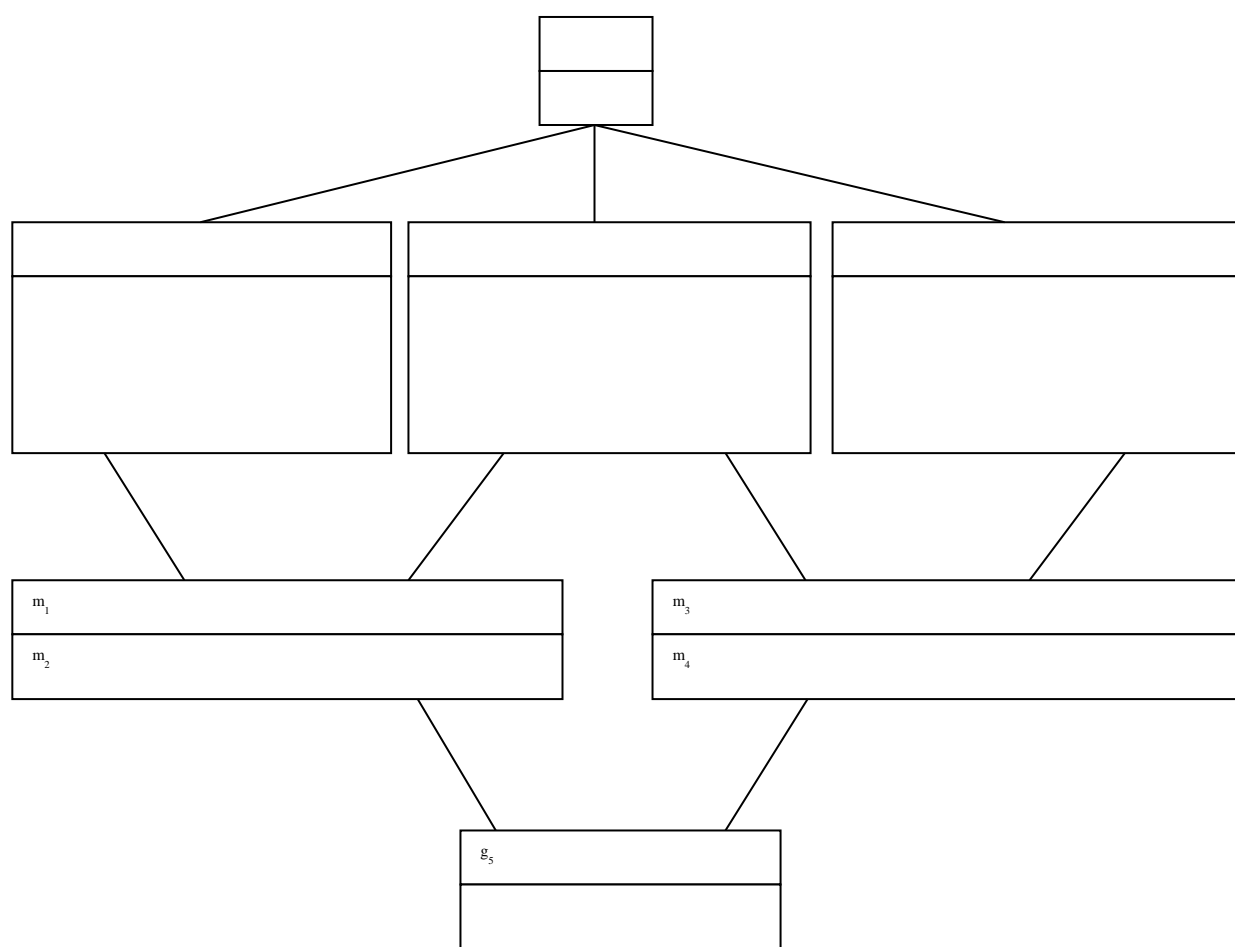
<sup>205</sup> Figure 9 de S.Reusch, élaboré d'après les informations tirés du Sommaire (pp. xv/xvi, articles 101. Zéro à 111. Dix) et du Corpus (p. 59, article 100. Nombre) du Thésaurus Larousse (1992).

<sup>206</sup> Cf. Priss (2005), pp. 155/156.

<sup>207</sup> Cf. Priss (2005), pp. 155/156.

Dans ce cas, le *contexte formel* est visualisé par un treillis, qui montre la relation entre un *concept* (terme générique) et les différentes *notions* (termes spécifiques) attachées à celui-ci.

Un tel treillis ne représente pas explicitement l'*extension* ou l'*intension*, car ils sont compris dans l'origine conceptuelle des différents termes.<sup>208</sup> La figure 10 explicite, par exemple, les différentes notions de '*concept*' que l'on trouve dans le Thésaurus Larousse.<sup>209</sup>



**Figure 10 : Les différentes notions de « concept » dans le Thésaurus Larousse**<sup>210</sup>

Mais comment créer un treillis qui rassemble les termes qui apparaissent dans différentes entrées de dictionnaire ? La figure 11 propose un schéma pour expliquer le procédé. À la

<sup>208</sup> Cf. Wille (2005), pp. 6/7 ; cf. Old (2004), pp. 237/238.

<sup>209</sup> Cf. Sedelow (1993), pp. 320/321.

<sup>210</sup> Figure 10 de S.Reusch. Élaboré d'après les informations trouvés dans l'index pour l'item « concept » (p. 728) et dans les articles 1. Existence (p.1), 421. Idée (pp. 302/303) et 478. Philosophie (p. 336) du Thésaurus Larousse (1992). Le schéma a été crée d'après les exemples consultés dans Sedelow (1996), p. 7 et d'après la consultation d'un treillis crée à partir de Roget's International Thesaurus pour le terme « concepts », sur le site internet « Formal Concept Analysis » de Uta Priss : <http://www.ketlab.org.uk/roget.html>.

recherche d'un mot, l'utilisateur consulte l'index du thésaurus - qui paraît sur les dernières pages du dictionnaire. L'organisation de l'index montre qu'un thésaurus classe les mots selon les idées qu'ils expriment.<sup>211</sup> Un mot figure ainsi dans différentes entrées dictionnairique, qui présentent les différentes *notions* de concepts. Une fois terminée la recherche dans l'index, l'utilisateur procède à la consultation du corpus. Il étudie toutes les entrées dans lesquelles figure le mot recherché et prend note de la *classe conceptuelle* et des mots du *voisinage sémantique*. Ensuite, il crée un treillis, en utilisant les *mots* du voisinage sémantique<sup>212</sup> comme *objets* et les *classes conceptuelles*<sup>213</sup> comme des *attributs*. Ce treillis permet à l'utilisateur l'étude des différentes *notions conceptuelles*.<sup>214</sup>

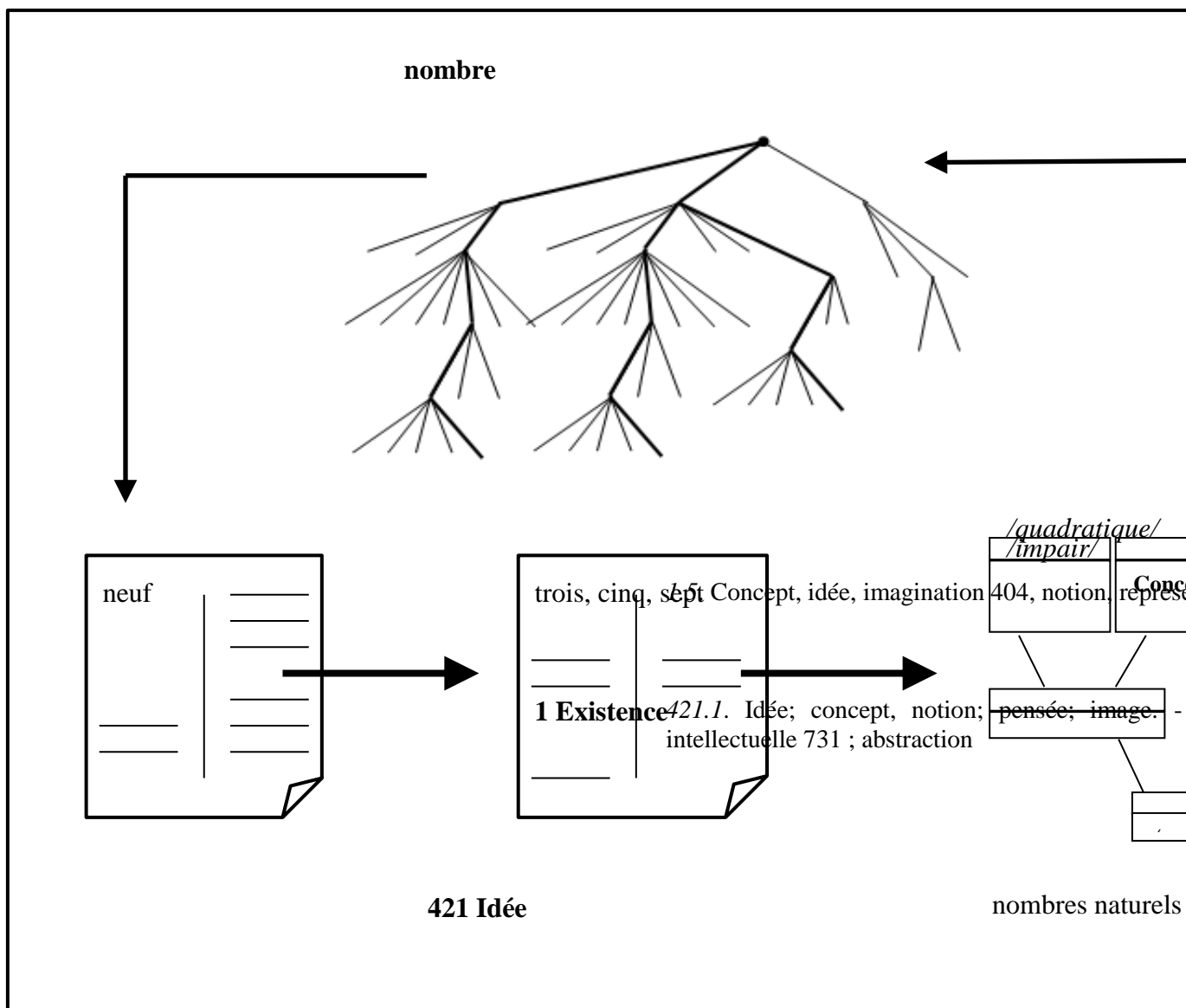
---

<sup>211</sup> Cf. Thésaurus Larousse (1992), Préface, p. v.

<sup>212</sup> Les mots du 'voisinage sémantique', ce sont dans notre cas des termes issus du même paragraphe que le mot de recherche dans une entrée de dictionnaire.

<sup>213</sup> Les 'classes conceptuelles', ce sont dans notre cas les entrées de dictionnaire qui dénotent les différents concepts.

<sup>214</sup> Cf. Priss (2005), pp. 155/156.



**Figure 11: L'analyse formelle des notions d'un concept basé sur la répartition des mots**<sup>215</sup>

L'*analyse formelle des concepts* est une méthode d'analyse des données et de la représentation des connaissances qui peut fournir un outil précieux pour l'étude de l'organisation conceptuelle d'un thésaurus. Les thésauri - comme le Thésaurus Larousse - contiennent en général des composants hiérarchiques, telle que la relation d'hyperonymie/hyponymie ou des hiérarchies de type. Les treillis - en tant qu'ensembles ordonnés - sont une représentation naturelle des hiérarchies et des systèmes de classification.<sup>216</sup>

<sup>215</sup> Figure 11 de S.Reusch. Élaboré d'après les expériences faits lors de la création du treillis « concept ».

<sup>216</sup> Cf. Old (2004), p. 236 ; cf. Priss (2005), pp. 155- 157.

Chaque thésaurus peut être représenté ou analysé par un treillis.<sup>217</sup> La formalisation d'un thésaurus avec la méthode AFC permet en conséquence une analyse linguistique, cognitive et anthropologique de la structure et des connaissances qui sont encodées dans ce type de dictionnaire.<sup>218</sup>

Par ses fondements mathématiques, l'*analyse formelle des concepts* peut aussi fournir le cadre pour une nouvelle définition de la *similarité sémantique* et de l'*analogie*, deux relations importantes pour le classement conceptuel et la structuration des connaissances. Nous utiliserons donc des fondements mathématiques pour procéder à une définition formelle de la similarité sémantique et de l'analogie dans le chapitre suivant. La définition - présentée de façon contrastive - vise à établir une distinction entre les deux notions, qui nous semble essentielle à la compréhension de l'organisation conceptuelle d'un thésaurus. La distinction ainsi établie soulignera que la similarité sémantique se situe au sein d'un thésaurus du côté de l'*organisation explicite*, alors que l'analogie se situe du côté de l'*organisation implicite* des connaissances.

### 3.2 Définition de la similarité sémantique et de l'analogie

Un thésaurus est constitué d'un ensemble partiellement ordonné de *termes*. Les *descripteurs* (ou *termes préférés*) forment les unités de base - sémantiques et cognitives - qui regroupent les *termes non-préférentiels* selon leur sens ou leur champs de référence (cf. chapitre 1.3.2 et 1.3.3).<sup>219</sup> L'ensemble des termes représente des *objets* ou des *types d'objets*. Ces *objets*, ainsi que leurs *termes* connexes, peuvent être caractérisées par des *attributs* (cf. chapitre 3.1.1).<sup>220</sup>

Le Thésaurus Larousse présente le terme préféré '*maison*'. Associés à ce terme générique on trouve des termes synonymes (ou termes non-préférentiels), tels que '*demeure*' ou '*domicile*'. Et, des termes spécifiques tels que '*cabanon*', '*ferme*', '*villa*', '*château*' ou '*château fort*'.<sup>221</sup> Tous ces termes décrivent des objets du monde réel dans un contexte de */bâtiment destiné à servir d'habitation à l'homme/*.<sup>222</sup> Un tel contexte, nous permet de comparer les objets et de caractériser ainsi des objets plus ou moins similaires à l'intérieur

---

<sup>217</sup> Cf. Priss (2005), p. 149.

<sup>218</sup> Cf. Priss (2005), p. 157.

<sup>219</sup> Cf. Reisinger (1977), p.140 ; cf. Taddei Elmi (1981), p. 847 ; cf. Bianucci (1992), p.60 ; cf. Renard et al. (2009), p. 88 ; cf. Reusch (2011b), p. 938.

<sup>220</sup> Cf. Wille (2005), p. 5 ; Reusch (2011b), p. 938.

<sup>221</sup> Termes tirés du Thésaurus Larousse (1992), article 848 Maison, p. 600.

<sup>222</sup> Cf. Le Nouveau Petit Robert de la langue française (2007), Article « maison », p. 1509 .

d'une même *classe d'objets*. Dans le contexte de '*maison*', nous pouvons constater, par exemple, que '*château-fort*' et '*château*' sont des **objets similaires**. Ils présentent des propriétés partagées, concernant leur taille, leur utilisation comme bâtiment, mais présentent aussi des propriétés non-partagées qui déterminent leur caractéristique spécifique. Une propriété non-partagée concerne par exemple le mode de construction : Un '*château-fort*' est fortifié et un '*château*' est non-fortifié (cf. figure 12).

Nous pouvons faire également la comparaison des termes qui figurent dans différents *contextes* et de leurs objets qui apparaissent dans différentes *classes d'objets*. Nous pouvons comparer, par exemple une '*tige de blé*' à une '*tour de télévision*'.<sup>223</sup> Le terme non-préférentiel '*tige*' figure dans le Thésaurus Larousse dans un article '*fleurs*' et appartient à la classe des '*plantes*'. Le terme non-préférentiel '*tour*' figure dans un article '*architecture*' et appartient à la classe des '*bâtiments*'. Les deux objets peuvent être considérés comme 'similaires' en ce qui concerne leur forme et leur usage premier. Leur forme est celle d'un cylindre avec un sommet élargi. La partie ultérieure du cylindre doit tenir une charge lourde, à un niveau très élevé au-dessus du sol (l'épi ou le transmetteur). Néanmoins, ils sont inscrits dans deux contextes différents par leur appartenance à deux classes d'objets distinctes. Nous définissons ainsi, des objets 'similaires' de contextes différents comme des **objets analogues** (cf. figure 13).

La théorie de l'analyse formelle de concepts peut être appliquée à un thésaurus. Dans cette perspective nous considérons : Le *descripteur* (ou *terme préféré*) au sein d'un thésaurus désigne un concept. Un concept est défini comme une « représentation symbolique (de la réalité) qui a une signification générale qui convient à une série d'objets concrets possédant des propriétés communes ».<sup>224</sup> De plus, un *contexte* doit être défini pour les *objets* et leurs *propriétés* afférentes.

Un contexte K est un triplet  $K(G,M,I)$ <sup>225</sup>. Il est constitué d'un ensemble d'objets G, d'un ensemble de propriétés M et d'une relation  $I \subset G, M$  qui affecte les objets  $g \in G$  à des propriétés  $m \in M$ .<sup>226</sup>

<sup>223</sup> Cf. Thésaurus Larousse (1992), articles 288.3 Fleurs (pp. 180-185) et 777.11 Architecture (pp. 535-537).

<sup>224</sup> Cit d'après Dubois et al. (1994), p.107.

<sup>225</sup> Cf. Wille (2005), p.2.

La notation du contexte formel  $K(G,M,I)$  est dérivée de la terminologie officielle en allemand : Kontext (K), Gegenstand (G), Merkmal (M), Inzidenz (I), qui pourrait être traduit en français par C(O,A,I) : Contexte (C), Objet (O), Attribut (A), Incidence (I).

<sup>226</sup> Cf. Wille (2005), pp. 2/3 ; Koester (2006), p. 13 ; Reusch (2011b), p. 938.

Pour un sous-ensemble  $G_1$  des objets de  $G$ ,  $G_1 \subset G$ , nous définissons un sous-ensemble dérivé  $D(G_1)$  de la totalité des propriétés comprises dans  $G_1$ , par :

$$D(G_1) = \{ m \in M \mid (g,m) \in I \text{ pour tous les } g \in G_1 \} .^{227}$$

Pour un sous-ensemble  $M_1$  des propriétés de  $M$ ,  $M_1 \subset M$ , nous définissons un sous-ensemble dérivé  $F(M_1)$  de la totalité des objets correspondants à des attributs dans  $M_1$ , par :

$$F(M_1) = \{ g \in G \mid (g,m) \in I \text{ pour tous les } m \in M_1 \} .^{228}$$

$C(G_1, M_1)$  est un concept de  $K(G,M,I)$ , si :

$$G_1 \subseteq G$$

$$M_1 \subseteq M$$

$$D(G_1) = M_1$$

$$F(M_1) = G_1 .^{229}$$

Considérant l'ensemble  $CON$  de tous les concepts pour le contexte  $K$ , une relation d'ordre peut être introduite pour l'ensemble de concepts de la façon suivante :

$$C_1 \leq C_2 \text{ si } G_1 \subseteq G_2,$$

où  $C_1 (G_1, M_1)$  et  $C_2 (G_2, M_2)$  sont des concepts arbitraires.

$(CON, \leq)$  est un treillis.<sup>230</sup>

Le concept fondamental sur lequel repose un treillis est celui de la *relation d'ordre*, un concept beaucoup discuté dans le contexte du thésaurus. La compréhension de la structure explicite d'un thésaurus doit partir d'une compréhension du principe organisateur adopté par les concepteurs. En outre, une compréhension approfondie des relations qui traitent les structures implicites - comme la *similarité sémantique* doit être développé. Ainsi, on peut étudier le thésaurus non seulement comme une structure classificatoire, mais encore

---

<sup>227</sup> Cf. Reusch (2011b), p. 938.

<sup>228</sup> Cf. Reusch (2011b), p. 938.

<sup>229</sup> Cf. Reusch (2011b), p. 938.

<sup>230</sup> Cf. Reusch (2011b), p. 938.



comme un réseau des relations. Nous proposons donc de poursuivre le chapitre avec une définition des notions de *similarité sémantique* et d'*analogie*.

### 3.2.1 Définition de la similarité sémantique

Étant donné un contexte  $K(G,M,I)$  et des objets  $g$  et  $f$  de  $G$  avec l'ensemble des attributs  $ATT(g)$  et  $ATT(f)$ ,

$$ATT(g) = \{ m \in M \mid (g,m) \in I \} \subset M \text{ et}$$

$$ATT(f) = \{ n \in M \mid (f,n) \in I \} \subset M. \text{ }^{231}$$

Les objets  $g$  et  $f$  sont **similaires**, si et seulement si:

1. L'ensemble des attributs CATT de  $g$  et  $f$  n'est pas vide :

$$CATT(g,f) = ATT(g) \cap ATT(f) \neq \emptyset$$

2. L'ensemble des attributs communs CATT est plus grand que l'ensemble des attributs différents DATT:

$$DATT(g, f) = (ATT(g) - ATT(f)) \cup (ATT(f) - ATT(g))$$

$$|DATT(g, f)| \leq |CATT(g,f)| \text{ }^{232}$$

La figure 12 explicite la relation de similarité sémantique entre les termes '*château*' et '*château-fort*'. Les deux termes se situent dans la même classe d'objets et dépendent, dans le Thésaurus Larousse, d'un même terme générique '*maison*'. La *similarité sémantique* entre '*château*' ( $f$ ) et '*château-fort*' ( $g$ ) dépend de deux ensembles de propriétés : Un ensemble des propriétés partagées (désigné  $CATT(g,f)$ ) et un ensemble des propriétés non-

<sup>231</sup> Cf. Reusch (2011b), p. 938.

<sup>232</sup> Cf. Reusch (2011b), p. 938 ; cf. Monneret (2004), pp. 11/12.

DATT ( $g,f$ ) : notation pour « set of different attributes of objects  $g$  and  $f$  »

CATT ( $g,f$ ) : notation pour « set of common attributes of objects  $g$  and  $f$  »



### Figure 12: La similarité sémantique entre château et château fort <sup>236</sup>

La similarité entre les objets  $g$  et  $f$  est une relation symétrique. D'autres conditions peuvent être ajoutées pour des types spécifiques de similarité. En outre, on pourrait songer à étendre la simple comparaison quantitative du nombre d'attributs communs ( $CATT(g,f)$ ) et d'attributs différents ( $DATT(g,f)$ ) et à la limiter, en quelque sorte, à des attributs qui ont une valeur significative pour la comparaison de deux objets. <sup>237</sup> Dans cette perspective, le grand nombre d'attributs négligeables ne pourrait jamais dépasser le petit nombre d'attributs d'importance significative, les dits *attributs de base*. L'équation pour la définition de la similarité sémantique, présenté dans ce chapitre, peut être améliorée pour prendre en considération la *valeur significative* des attributs pour une comparaison de deux objets, exprimé par un *poids*. Une telle spécification pourrait être l'enjeu d'un futur travail.

#### 3.2.2 Définition de l'analogie

Étant donné un contexte  $K(G_1, M_1, I_1)$  et un contexte  $K(G_2, M_2, I_2)$  et  $G_1 \cap G_2 = \emptyset$  et

$$M_1 \cap M_2 \neq \emptyset$$

Les objets  $g_1 \in G_1$  et  $g_2 \in G_2$  avec l'ensemble des attributs  $ATT(g_1)$  et  $ATT(g_2)$ ,

$$ATT(g_1) = \{ m_1 \in M_1 \mid (g_1, m_1) \in I_1 \} \subset M_1 \text{ et}$$

$$ATT(g_2) = \{ m_2 \in M_2 \mid (g_2, m_2) \in I_2 \} \subset M_2. \quad ^{238}$$

Les objets  $g_1$  et  $g_2$  sont **analogues**, si et seulement si:

1. L'ensemble des attributs communs  $CATT$  de  $g$  et  $g_2$  n'est pas vide :

$$CATT(g_1, g_2) = ATT(g_1) \cap ATT(g_2) \neq \emptyset$$

3. L'ensemble des attributs communs  $CATT$  est plus grand que l'ensemble des attributs différents  $DATT$ :

---

<sup>236</sup> Figure 12 de S.Reusch. Élaboré d'après les informations trouvés dans le Thésaurus Larousse (1992), article 848. Maison, pp. 600-602.

<sup>237</sup> Cf. Reusch (2011b), p. 938.

<sup>238</sup> Cf. Reusch (2011b), p. 938.

$$\text{DATT}(g_1, g_2) = (\text{ATT}(g_1) - \text{ATT}(g_2)) \cup (\text{ATT}(g_2) - \text{ATT}(g_1))$$

$$|\text{DATT}(g_1, g_2)| \leq |\text{CATT}(g_1, g_2)|^{239}$$

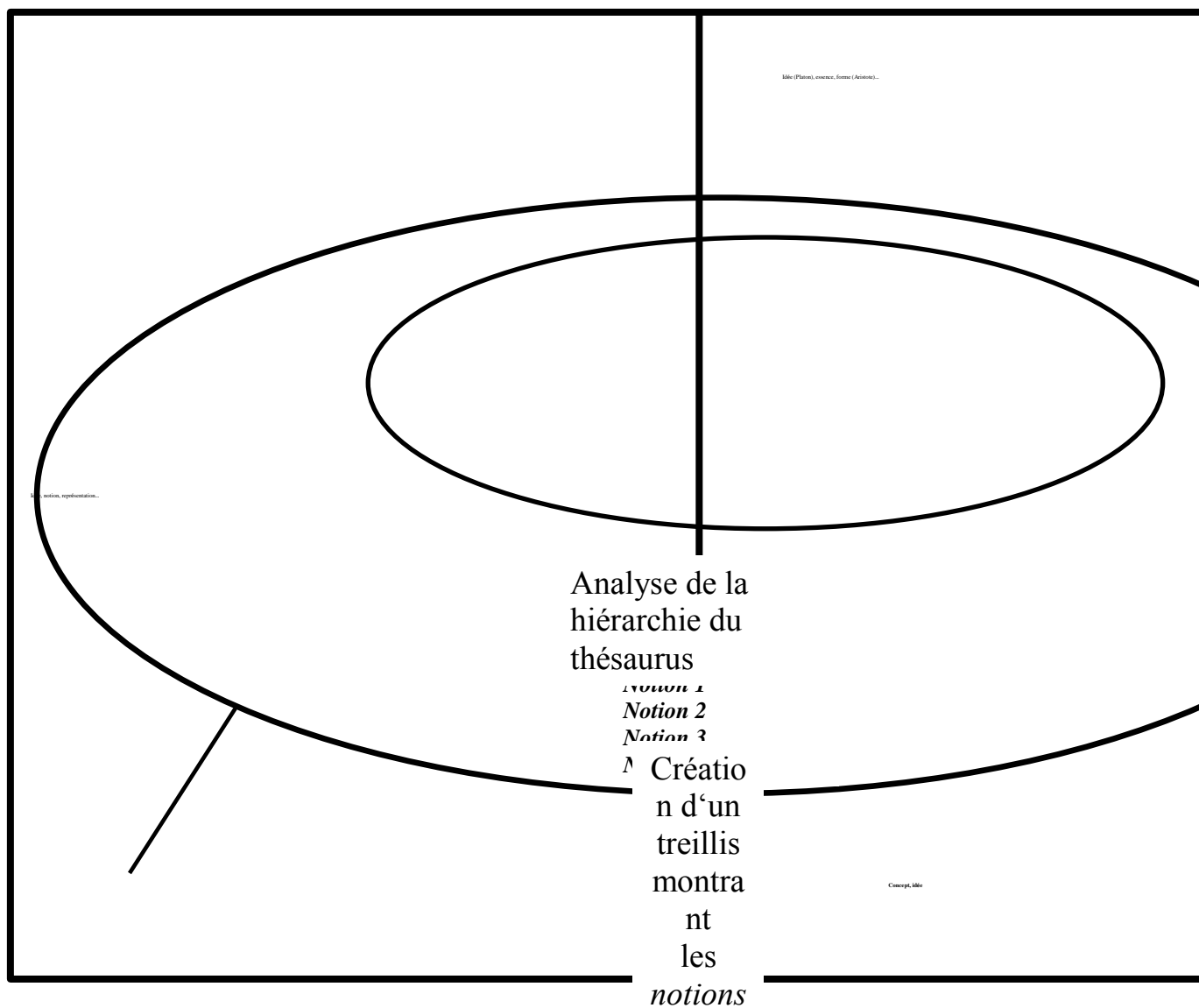
La figure 13 explicite la relation d'analogie entre les termes '*tige de blé*' et '*tour de télévision*'. Les deux termes se situent dans deux classes d'objets différents : La tige appartient à la classe des *plantes*, la tour appartient à la classe des *bâtiments*. Classés au sein du Thésaurus Larousse, les deux termes sont régis par différents termes génériques : '*Tige de blé*' dépend du terme générique '*fleur*', '*tour de télévision*' dépend du terme générique '*architecture*'. Une appartenance de classe divergente est le facteur déterminant pour l'établissement d'une analogie entre termes. Le schéma montre que la relation d'analogie entre '*tige de blé*' ( $g_1$ ) et '*tour de télévision*' ( $g_2$ ) dépend de deux ensembles de propriétés : Un ensemble des propriétés partagées (désigné  $\text{CATT}(g_1, g_2)$ ) et un ensemble des propriétés non-partagées (désigné  $\text{DATT}(g_1, g_2)$ ).<sup>240</sup> L'ensemble  $\text{DATT}(g_1, g_2)$  contient toutes les propriétés génériques qu'héritent '*tige de blé*' et '*tour de télévision*' de leurs termes génériques '*fleur*' et '*architecture*' (comme */matière organique/* et */matière inorganique/*), ainsi que toutes les propriétés spécifiques qui caractérisent les objets dans le cadre de leur classe d'objets. L'ensemble  $\text{CATT}(g_1, g_2)$  contient toutes les propriétés sur lesquelles repose l'analogie. Il s'agit des propriétés génériques (comme */stationnaire/*) ou spécifiques (comme */de forme cylindrique/*) qui caractérisent deux objets de classes différents comme 'similaires' en vue de certaines caractéristiques. Deux objets de classes différents ont beaucoup de 'similarités', si leur ensemble des propriétés partagées  $\text{CATT}(g_1, g_2)$  est plus grand que l'ensemble des propriétés non-partagées  $\text{DATT}(g_1, g_2)$ .<sup>241</sup>

<sup>239</sup> Cf. Reusch (2011b), p. 938 ; cf. Monneret (2004), pp. 11/12.

L'évaluation des propriétés communes et des propriétés différentes, ainsi que leur pondération (selon la définition), dépend de l'expansion de la représentation (mesurée par exemple en *slots* disponibles, cf. chapitre 3.3.2.2), le choix des représentants, etc. . En conséquence, notre définition de l'analogie ne peut pas prétendre à l'universalité.

<sup>240</sup> Cf. Reusch (2011b), p. 938 ; cf. Kunzmann (1998), pp. 87/88 ; cf. Monneret (2004), pp. 11/12.

<sup>241</sup> Cf. Reusch (2011b), p. 938 ; cf. Kunzmann (1998), pp. 87/88 ; cf. Monneret (2004), pp. 11/12.



**Figure 13: L'analogie entre tige de blé et tour de télé**

L'analogie entre les objets  $g_1$  et  $g_2$  est une relation *symétrique*. d'autres conditions peuvent être ajoutées pour des types d'analogie spécifiques. Par contraste avec l'évaluation de la *similarité sémantique* entre mots, nous comparons lors de l'évaluation de l'*analogie* les objets issus de deux contextes différents, comme par exemple la *tour de télévision* et la *tige de blé*. Néanmoins, nous avons besoin des attributs communs - sinon une comparaison entre deux objets serait impossible. De plus, la simple comparaison quantitative du nombre d'attributs communs et du nombre d'attributs différents doit être étendue et doit être limitée aux attributs de valeur significative. Dans l'exemple, la tour de télévision et la tige de blé ont beaucoup d'attributs en commun, qui concernent la forme, la proportion, l'utilisation et

<sup>242</sup> Figure 13 de S.Reusch. Élaboré d'après les informations trouvés dans le Thésaurus Larousse (1992), articles 288. Fleurs (pp. 180-185) et 777. Architecture (pp. 535/536).

même les sous-structures spécifiques pour la stabilisation. Mais, il existe également des attributs différents, qui concernent le matériel, la taille réelle etc. Ainsi, nous avons besoin d'une mesure fiable pour pouvoir comparer la nature de tels objets. La relation d'analogie peut aussi être établie sur la base d'un seul attribut partagé par deux objets - ceci est le cas souvent lors qu'on compare deux objets de nature différente lors du *raisonnement par analogie* ou lors de la création d'une *métaphore*.

### 3.2.3 Définition de la métaphore

La *métaphore* concerne un cas spécifique de l'analogie. « La métaphore est le transport à une chose d'un nom qui désigne une autre » [Aristote, Poétique, 20].<sup>243</sup> La métaphore a une *fonction cognitive* qui permet de comparer deux objets qui figurent dans des domaines sémantiques très éloignés, comme par exemple un objet figurant dans la classe des *humains* et un objet figurant dans la classe des *animaux*.<sup>244</sup> Ainsi, on utilise la métaphore « *Cet enfant est une tortue.* » pour exprimer qu'un enfant marche lentement. Dans ce cas, nous sommes uniquement intéressé par un seul attribut (*/lent/*) qui est typique pour une tortue et le transfère à l'enfant (cf. figure 14).

Étant donné un contexte  $K(G_1, M_1, I_1)$  et un contexte  $K(G_2, M_2, I_2)$  et  $G_1 \cap G_2 = \emptyset$  et  $M_1 \cap M_2 \neq \emptyset$

et les objets  $g_1 \in G_1$  et  $g_2 \in G_2$  avec l'ensemble des attributs  $ATT(g_1)$  et  $ATT(g_2)$ ,

$ATT(g_1) = \{ m_1 \in M_1 \mid (g_1, m_1) \in I_1 \} \subset M_1$  et

$ATT(g_2) = \{ m_2 \in M_2 \mid (g_2, m_2) \in I_2 \} \subset M_2$ .<sup>245</sup>

Un objet  $g_1$  (*tortue*) avec un attribut spécifique  $m_1$  (*lent*) construit une métaphore pour un objet  $g_2$  (*enfant*), si et seulement si :

$m_1$  ne présente pas un attribut caractéristique de  $M_2$  (attributs caractéristiques pour l'enfant normal).<sup>246</sup>

La figure 14 donne un schéma explicatif de la création de la métaphore « *Cet enfant est une tortue* ».

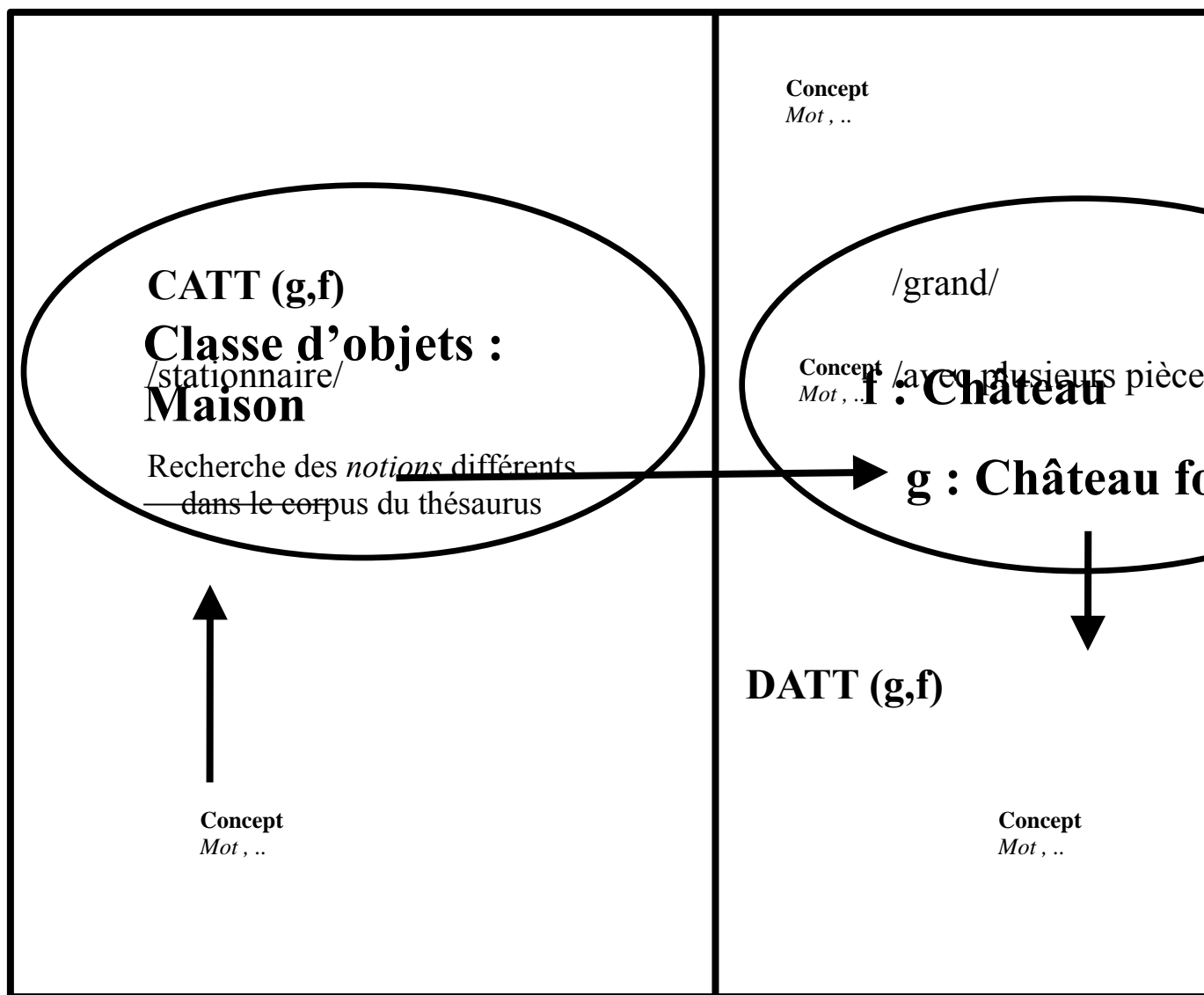
---

<sup>243</sup> Cité d'après Tijus (2003), p. 37 ; cf. aussi Itkonen (2005), p. 37.

<sup>244</sup> Cf. Tijus (2003), p. 38 ; cf. Taylor (2009), p. 135 ; cf. Itkonen (2005), pp. 35-44 ; cf. Charconnet (2003), p. 48 ; cf. Old /Priss (2001), pp. 99-104.

<sup>245</sup> Cf. Reusch (2011b), p. 938.

<sup>246</sup> Cf. Reusch (2011b), p. 938.



**Figure 14: Métaphore créée sur la base des deux termes ‘tortue’ et ‘enfant’** <sup>247</sup>

L'*analyse formelle des concepts* sert de fondement pour l'étude des relations de similarité sémantique et d'analogie. La méthode fournit un *contexte formel*  $K(G,M,I)$  dans lequel peut être évaluée la proximité sémantique entre deux objets figurant dans une ou plusieurs classes d'objets. La définition de la similarité sémantique et de l'analogie qui se déduit de cette analyse est importante pour la compréhension du classement sémantique opéré dans un thésaurus. La *similarité sémantique* est donnée lorsque deux objets, qui figurent dans une même classe d'objets, sont sémantiquement proches. L'*analogie* est donnée lorsque deux objets, qui figurent dans deux classes d'objets différentes, sont rapprochés à cause d'une proximité sémantique ponctuelle. L'*appartenance de classe* est dans notre définition

<sup>247</sup> Figure 14 de S.Reusch. Élaboré d'après les informations tirés de Tijus (2003), p. 38, de Itkonen (2005), pp. 35-44 et de Charconnet (2003), p. 48.



la clé pour une distinction entre les deux types de relations. Dans cette perspective, la *similarité sémantique* est située au côté de l'*organisation hiérarchique (explicite)* du thésaurus, qui organise les objets du monde réel en catégories conceptuelles. L'*analogie* est située au côté de l'*organisation sémantique (implicite)* du thésaurus, qui permet de rapprocher les objets de diverses catégories conceptuelles lors de la consultation. L'*appartenance de classe* situe l'évaluation de la similarité sémantique du côté *statique* : Dans un thésaurus elle fait l'objet d'une simple consultation d'articles préexistants. L'évaluation de l'analogie se situe du côté *dynamique* : Dans un thésaurus elle fait l'objet d'une consultation active de divers articles, entre lesquelles sont instaurés des liens sémantiques. Une compréhension approfondie des relations de similarité sémantique et d'analogie, nous permet de mieux comprendre la différence entre l'*organisation explicite* et l'*organisation implicite* d'un thésaurus. L'*analyse formelle des concepts* ne permet pas uniquement de définir similarité sémantique et analogie, mais aussi d'étudier en détail sur quels liens repose la relation établie - ceci sera le sujet du prochain chapitre.

### 3.3 Analyse formelle de la similarité sémantique et de l'analogie

L'*analyse formelle des concepts* est une méthode précieuse pour l'analyse globale du réseau des relations qu'établie dans un thésaurus et pour l'analyse détaillée du lien de similarité sémantique et d'analogie qui s'instaure entre deux objets (cf. chapitre 3.1 et 3.2).

#### 3.3.1 Analyse formelle de la similarité sémantique

L'*analyse formelle des concepts* nous permet d'instaurer un contexte formel  $K(G,M,I)$  pour 'maison', un terme générique du Thésaurus Larousse. Le contexte  $K$  se constitue d'un ensemble d'objets  $G$  qui contient les termes spécifiques de l'article 848 *Maison*, tel que 'cabanon', 'ferme', 'villa' et 'château-fort'.<sup>248</sup> Le contexte  $K$  associe aux termes spécifiques un ensemble des propriétés  $M$ , qui contient des attributs tel que 'stationnaire', 'construit en bois' ou 'habité'. Dans le contexte formel  $K$ , l'ensemble d'objets  $G$  et l'ensemble de propriétés  $M$  sont liés par une relation d'incidence  $I \subset G,M$ .<sup>249</sup> La relation d'incidence permet d'étudier les termes spécifiques en vue de leurs propriétés partagées et

<sup>248</sup> Cf. Wille (2005), pp. 2/3 ; Koester (2006), p. 13 ; Reusch (2011b), p. 938.

<sup>249</sup> Cf. Wille (2005), pp. 2/3 ; Koester (2006), p. 13 ; Reusch (2011b), p. 938.

non-partagées (cf. chapitre 3.2.1).<sup>250</sup> Le tableau 16 montre un tableau croisé qui présente le contexte formel du concept '*maison*'.

**Tableau 16: Étude de similarité sémantique des termes spécifiques de « maison »**<sup>251</sup>

attributs (m <sub>n</sub> ) objets (g <sub>n</sub> )	m <sub>1</sub> : /stationnaire/	m <sub>2</sub> : /construit en bois/	m <sub>3</sub> : /habité/	m <sub>4</sub> : /fortifié/	m <sub>5</sub> : /attaché à une surface agricole/
g <sub>1</sub> : cabanon	x	x			
g <sub>2</sub> : ferme	x		x		x
g <sub>3</sub> : villa	x		x		
g <sub>4</sub> : château	x		(x)		
g <sub>5</sub> : château fort	x			x	

L'*analyse formelle* du concept '*maison*' montre l'*extension*, les objets du monde réel *cabanon*, *ferme*, *villa*, *château* et *château-fort* qui présentent les différentes réalisations du concept '*maison*' et montre l'*intension*, les attributs que possède une '*maison*' dans ses différentes manifestations. Une '*maison*' peut être /stationnaire/, /construit en bois/, /habité/, /fortifié/ ou peut être /attaché à une surface agricole/.

L'*analyse formelle* du concept '*maison*' montre aussi les *incidences objet-attribut*, qui permet de rapprocher différents *types d'objet* en vue de leurs propriétés partagées. L'attribut /stationnaire/ forme une incidence pour tous les objets concernés. La rangée de croix continue (cf. colonne 1) indique que /stationnaire/ est un *attribut de base*, un *attribut générique* que tous les objets héritent de '*maison*'. Les attributs /construit en bois/, /fortifié/ et /attaché à une surface agricole/ forment une incidence pour un objet unique. La croix individuelle (cf. colonne 2, 3 et 5) indique qu'il s'agit des *attributs spécifiques* qui font la particularité d'un type d'objet. L'utilisation particulière d'un type de maison décide sur ses propriétés spécifiques. Le '*cabanon*', comme pavillon de chasse utilisé en vacances, a pour particularité d'être /construit en bois/, le '*château fort*', comme défense

<sup>250</sup> Termes tirés du Trésor Larousse (1992), article 848. Maison, pp. 600-602.

<sup>251</sup> **Tableau 16** de S.Reusch. Élaboré d'après les informations trouvés dans le Trésor Larousse (1992), article 848. Maison, pp. 600-602.

militaire, d'être /fortifié/ et la 'ferme', comme site de production agricole, d'être /attaché à une surface agricole/. L'attribut /habité/ se situe entre les deux types d'incidences. L'attribut /habité/ est un attribut générique de 'maison' qui est partagée par les objets 'villa', 'ferme' et 'château'. En raison d'évolutions - historiques ou objectales - 'château fort' et 'cabanon' ne partagent pas cette propriété.

L'analyse formelle de concepts repose sur une *représentation* d'un contexte formel  $K(G,M,I)$ , son institution par un choix d'attributs et d'objets est déterminant pour le résultat de l'analyse de similarité sémantique entre *deux* objets.<sup>252</sup> Le contexte formel peut montrer que deux termes, considérés comme sémantiquement proches, peuvent avoir beaucoup de propriétés non-partagées (exprimé par les *non-incidences*) : Les termes 'château' et 'château fort' ont une seule incidence partagée, celle d'être /stationnaire/ (cf. tableau 16). Cependant, cette incidence unique prime sur la masse des autres, car elle indique que la similarité sémantique entre les termes repose sur une propriété générique, héritée du concept « maison ». L'analyse de concepts amène la preuve que les propriétés génériques et les propriétés spécifiques ne partagent pas le même statut dans la détermination de la similarité sémantique entre termes.

### 3.3.2 Analyse formelle de l'analogie

L'*analyse formelle des concepts* permet aussi de d'étudier la relation d'*analogie* qui peut être instaurée entre termes de domaines sémantiques différents. La méthode nous permet d'instaurer un contexte formel  $K(G,M,I)$  pour 'avion', 'montgolfière' et 'oie', trois termes spécifiques du Thésaurus Larousse.<sup>253</sup> Les trois termes constituent l'ensemble d'objet G, qui dépend de deux domaines sémantiques différents : Dans le Thésaurus Larousse, 'avion' et 'montgolfière' dépendent du terme générique 'moyens de transport' et 'oie' dépend du terme générique 'oiseaux'. Le contexte  $K$  associe aux termes spécifiques un ensemble des propriétés M, qui contient différents attributs que peuvent posséder les représentants de deux classes d'objets.<sup>254</sup> Ainsi, la méthode AFC permet d'étudier les termes en vue de propriétés partagées et non-partagées (cf. chapitre 3.2.2).<sup>255</sup> Le tableau 17 montre un tableau croisé qui présente le contexte formel des concepts 'moyens de transport' et 'oiseaux' qui peut être établi par alignement des objets 'avion', 'montgolfière' et 'oie'.

<sup>252</sup> Cf. Wille (2005), pp. 2/3 ; Koester (2006), p. 13 ; Reusch (2011b), p. 938.

<sup>253</sup> Cf. Wille (2005), pp. 2/3 ; Koester (2006), p. 13 ; Reusch (2011b), p. 938.

<sup>254</sup> Cf. Wille (2005), pp. 2/3 ; Koester (2006), p. 13 ; Reusch (2011b), p. 938.

<sup>255</sup> Termes tirés du Thésaurus Larousse (1992), article 297.7 Oiseaux (p.194) et 820.1 Transports par air (p.578).

**Tableau 17: Étude de la relation d'analogie entre 'avion', 'montgolfière' et 'oie'** <sup>256</sup>

attributs (m <sub>n</sub> ) objets (g <sub>n</sub> )	m <sub>1</sub> :/de matière inorganique /	m <sub>2</sub> :/peut changer de direction/	m <sub>3</sub> : /possède une envergure /	m <sub>4</sub> : /consomme de l'énergie/	m <sub>5</sub> : /a une masse max. au décollage/	m <sub>6</sub> : / vitesse de croisière fixée/	m <sub>7</sub> : / vol en nuées/	m <sub>8</sub> : /transporte des passagers /
g <sub>1</sub> : avion	x	x	x	x	x	x		x
g <sub>2</sub> : montgolfière	x			x	x			x
g <sub>3</sub> : oie		x	x	x	x	x	x	

L'analyse formelle des concepts 'moyens de transport' et 'oiseaux' montre comme *extension* les objets 'avion', 'montgolfière' et 'oie'. Elle montre comme *intension* différents attributs que peuvent posséder les représentants de deux classes d'objets : /être de matière inorganique/, /pouvoir changer de direction/, /posséder une envergure/, /consommer de l'énergie/, /avoir une masse maximale au décollage/, /avoir une vitesse de croisière fixe/, /voler en nuées/ et /transporter des passagers/. Une analogie entre 'moyens de transport' et 'oiseaux' peut être établie, si l'*extension* et l'*intension* des deux concepts est en convergence, exprimée par une *incidence objet-attribut*. La représentation montre qu'une analogie entre 'avion' et 'oie' peut être conçue sur un nombre abondant d'*incidences attribut-objet*. 'Avion' et 'oie' partagent cinq incidences, sur un chiffre total de huit incidences probables. Ils peuvent être considérés comme analogues concernant leur /prise de direction autonome/, leur /envergure/, leur /consommation d'énergie/, leur /masse maximale au décollage/, et leur /vitesse de croisière préfixée/. 'Montgolfière' et 'oie' partagent uniquement deux incidences, sur un chiffre total de huit incidences probables. Selon la représentation du contexte formel K(G,M,I), ils peuvent être considérés comme analogues concernant leur /consommation d'énergie/ et leur /masse maximale au décollage/. <sup>257</sup>

<sup>256</sup> Tableau 17 de S.Reusch. Élaboré d'après les informations trouvés dans le Thésaurus Larousse (1992), articles 297.7 Oiseaux (p.194) et 820.1 Transports par air (p.578).

<sup>257</sup> Cf. Reusch (2011b), p. 938.

L'analyse formelle des concepts montre que deux membres d'une classe d'objets, tels que 'avion' et 'montgolfière', peuvent être plus ou moins analogues à un troisième objet d'une classe différente, tel que 'oie'. L'appartenance de classe, ainsi que les attributs génériques hérités, tel que /consommation d'énergie/ et /masse maximale au décollage/, ne décident pas en dernière instance d'une proximité sémantique entre deux objets, mais de la masse de propriétés spécifiques partagées qui s'ajoute à celle-ci, lors de la comparaison de 'avion' et 'oie' par exemple, à voire la /prise de direction autonome/, l'/envergure/, et la /vitesse de croisière préfixée/ (cf. chapitre 3.2.2). Néanmoins, il ne faut pas oublier qu'une relation d'analogie peut aussi être établie sur *un seul attribut* en considération, lors du raisonnement par analogie (cf. chapitre 3.2.3). Ainsi, 'montgolfière', 'avion' et 'oie' pourraient être alignés par l'étude de leur /consommation d'énergie/.

L'analyse formelle des concepts nous a fourni, dans le chapitre présent, une possibilité de comparer des concepts à base d'une évaluation de l'incidence objet-attribut. La question qui se présente maintenant à nous, est celle de savoir de quelle *nature* sont les objets et leurs propriétés afférentes et, de quelle *origine* ils proviennent.<sup>258</sup> Pour pouvoir déterminer ceci, nous nous référons à la définition du concept que nous avons donné dans le chapitre 3.1.1. Dans celle-ci, nous avons fait remarquer qu'un concept est considéré comme l'*unité de base de la pensée* qui est formé dans un processus dynamique au sein des environnements sociaux et culturels.<sup>259</sup> L'homme élargit sa connaissance du monde (*phil. Weltwissen*) par *interaction* avec son environnement - il apprend par *intégration* et *adaptation* de concepts dans son appareil cognitif.<sup>260</sup> Les membres d'une communauté culturelle, partagent leurs *expériences* à l'aide des *concepts*. Par nécessité de partager leurs *expériences*, les membres partagent aussi les *moyens linguistiques* - en particulier une partie substantielle du lexique, mais aussi des options de formulation grammaticale - pour exprimer les concepts.<sup>261</sup> Un *concept* est donc accompagné d'un *système*, qui impose une *structure* et une *cohérence* sur certains aspects de l'expérience humaine.<sup>262</sup> L'analyse formelle de concepts rend transparent ce caractère systémique. L'analyse nécessite d'une *structure* auxiliaire pour pouvoir discerner l'*extension* et l'*intension* d'un concept, celle-ci est donnée par le *contexte formel*  $K(G,M)$ . Et elle requiert une *cohérence* dans la structure de données pour fournir une analyse communément acceptée, celle-ci est donnée par

<sup>258</sup> Cf. Konerding (1993), p.19.

<sup>259</sup> Cf. Wille (2005), p.2.

<sup>260</sup> Cf. Konerding (1993), p.11 et p.47.

<sup>261</sup> Cf. Konerding (1993), p.47 ; Fillmore (1985), pp. 231/232.

<sup>262</sup> Cf. Fillmore (1975), p.123.

l'*incidence attribut-objet* I, qui définit le *contexte formel* K comme *système d'attribution* K(G,M,I).

La grande faiblesse de l'*analyse formelle des concepts* est que sa référence à la réalité se limite au monde d'objets (*phil.* Objektwelt). Le système d'attribution K(G,M,I) se concentre uniquement sur l'*incidence objet-attribut* sans tenir compte de différents contextes culturels qui impliquent souvent différentes *expériences* sur la nature de concepts, et par conséquent sur les propriétés d'objets. Prenons l'exemple du concept « *maison* », un patrimoine de l'humanité. L'incidence '*maison*' - '*construit*' - '*en pierre*' est une expérience fixée par les communautés culturelles de l'Europe centrale. Elle se distingue des communautés culturelles de l'Afrique, où l'incidence '*maison*' - '*construit*' - '*en argile*' est une expérience communément acceptée.

La *Sémantique des cadres*, une théorie de linguistique cognitive que nous introduisons dans le chapitre suivant, prend en considération les expériences humaines, les faits culturels et sociaux, qui se cachent derrière le classement conceptuel d'un thésaurus. Pour cela, elle va élargir le contexte formel K(G,M,I) de l'*analyse formelle des concepts* de *valeurs* (all. *Werte*) que peut assumer une propriété d'objet. Ainsi, elle proposera un nouveau système d'attribution K(G,M,W,I)<sup>263</sup> et une méthode plus commode pour la représentation des connaissances à l'aide du *cadre conceptuel* même.

### 3.4 Sémantique des cadres

La *Sémantique des cadres* (angl. frame theory) est une théorie proposée par Charles J. Fillmore dans les années 1970.<sup>264</sup> L'idée de base est qu'on ne peut comprendre la signification d'un terme qu'en tenant compte du *cadre conceptuel* dans lequel il s'inscrit.<sup>265</sup> Dans la publication « An Alternative to Checklist Theories of Meaning » (1975) Fillmore définit un *cadre* (angl. frame) de manière suivante:

« Frames are "certain *schemata* or *frameworks of concepts or terms* which link together as a system, which impose structure or coherence on some aspects of

---

<sup>263</sup> Redéfinition du contexte formel K(G,M,I) par S. Reusch avec ajout de *valeurs pragmatiques* (all. *Werte, W*) que peuvent assumer de propriétés dans différents contextes culturels: K(G,M,W,I). Ce qui se déduit du nouveaux contexte formel est une triple *incidence objet-attribut-valeur*.

<sup>264</sup> Cf. Fillmore (1975), pp. 123-131 ; Konerding (1993), pp. 20/21 et pp. 43-49.

<sup>265</sup> Cf. Baker (2009), p. 32.

human experience and which may contain elements which are simultaneously parts of other such frameworks.” »<sup>266</sup>

Le *cadre* est, selon Fillmore, une structure d’information associée à un concept. Un cadre représente un concept (un individu, un objet ou une situation) dont les propriétés sont décrits à l’aide des *éléments cadres* (aussi appelés *slots*).<sup>267</sup>

La *Sémantique des cadres* permet d’étudier en détail le rapport objet-attribut. Le *cadre* représente un objet (maison) dont les attributs sont spécifiés à l’aide des *éléments cadres* (matériel etc.). Le rapport entre cadre et éléments cadre s’exprime donc par une variation possible des attributs d’objet. Les éléments cadre précisent donc aussi les *valeurs* possibles (pierre, argile etc.) que peut assumer une propriété d’objet.<sup>268</sup>

(Objet (Attribut : Valeur))

(maison (matériel : pierre))

En outre, les *éléments cadre* permettent d’associer deux concepts figurant dans deux *cadres* différents. L’association s’opère grâce à une *valeur de propriété* que partagent les deux objets en considération.<sup>269</sup>

(maison (matériel : pierre)) ↔ (montagne (matériel : pierre))

### 3.4.1 Le cadre - structure pour la comparaison de deux concepts

#### Exemple 1 :

Reprenons l’exemple de la *similarité sémantique* entre ‘*château*’ et ‘*château fort*’ pour illustrer la fonction principale du cadre sémantique entant que *structure d’informations* qu’illustre les propriétés partagées et non-partagées de deux objets en vue d’une *valeur*.

Les deux termes ‘*château*’ et ‘*château fort*’ figurent - comme nous l’avons déjà vu - dans une même entrée dictionnaire 848 *Maison* (cf. chapitre 3.3.1). Ils dépendent d’un même terme générique ‘*maison*’ qui inscrit les termes dans une *relation genre-espèce*.

<sup>266</sup> Cité d’après Kondering (1993), p. 21. Cf. aussi Fillmore (1975), p.123 (texte original).

<sup>267</sup> Cf. Archer (2009), pp. 37/38 .

<sup>268</sup> Cf. Mansour (1983), p. 5 ; cf. Archer (2009), pp. 37/38 .

<sup>269</sup> Cf. Mansour (1983), p. 6 ; cf. Archer (2009), pp. 37/38 .

L'appartenance au terme 'maison' implique l'héritage de tous ses propriétés inhérentes (aussi appelés *attributs génériques*), nommées dans cadre 1. Les attributs génériques définissent 'château' et 'château fort' comme une espèce de maison, exprimée par la relation AKO<sup>270</sup> qui relie le cadre 1 aux cadres 2 et 3. Le tableau 18 présente l'analyse du cadre sémantique de 'château' et 'château fort'.

**Tableau 18 : Analyse du cadre sémantique de 'château' et 'château fort'** <sup>271</sup>

{	CADRE 1 : Maison
	(f. maison)
	(CONSTRUCTION (BÂTIMENT))
	(EMPLACEMENT (FIXE))
	(CONSTRUCTEUR (HOMME))
	(FONCTION PRINCIPALE (HABITATION))
{	CADRE 2 : Château
	(m. château)
	(CHATEAU (AKO (MAISON)))
	(MATÉRIEL (EN PIERRE))
	(TAILLE (GRANDE))
	(INVESTISSEMENT FINANCIER (COÛTEUX))
	(MODE DE CONSTRUCTION (NON FORTIFIÉ))
	(FONCTION SPÉCIFIQUE (REPRÉSENTATION))
{	CADRE 3: Château fort
	(m. château fort)
	(CHÂTEAU FORT (AKO (MAISON)))
	(MATÉRIEL (EN PIERRE))
	(TAILLE (GRANDE))
	(INVESTISSEMENT FINANCIER (COÛTEUX))
	(MODE DE CONSTRUCTION (FORTIFIÉ))
	(FONCTION SPÉCIFIQUE (DÉFENSE))

La *Sémantique des cadres* est une méthode que démontre que deux termes, tels que 'château' et 'château fort', sont liés par une proximité sémantique, qui s'exprime par une ressemblance partielle. <sup>272</sup>

<sup>270</sup> Le signe AKO désigne la relation A-Kind-Of (fr. une espèce de).

<sup>271</sup> Tableau 18 de S.Reusch. Élaboré d'après un schéma proposé par Mansour (1983), p. 5 et d'après les informations trouvés dans le Thésaurus Larousse (1992), article 848. Maison, pp. 600-602.

<sup>272</sup> Cf. Monneret (2004), p.18.



Pendant l'analyse, le *cadre* fonctionne comme une structure d'informations qui indique par les *éléments cadre* identiques et non-identiques, les propriétés partagées et non-partagées d'objets - aussi en vue d'une *valeur* qui pourrait informer sur un contexte culturel particulier.<sup>273</sup>

L'analyse de la similarité sémantique implique toujours la notion d'*héritage des propriétés* d'un terme générique - dans notre cas, celui de '*maison*'. Le premier cadre informe sur les propriétés génériques que partagent deux objets similaires, en tant qu'*espèces* d'un même *genre*. '*Château*' et '*château fort*' partagent les propriétés d'être des *bâtiments*, donc des constructions à *emplacement fixe*, d'être *construit par l'homme* et surtout pour l'homme en tant qu'*habitation*.

L'analyse de la similarité implique aussi, toujours, une analyse de détail sur quels *éléments* (*attributs* et *valeurs*) repose une ressemblance partielle de deux objets qui se traduit linguistiquement par une '*proximité sémantique*'. '*Château*' et '*château fort*' partagent les mêmes propriétés d'objet. En tant que *bâtiment d'habitation*, ils se définissent par un *matériel*, une *taille* et un *mode* de construction, un *investissement financier* et une *fonction spécifique* qui s'attache au mode de construction. '*Château*' et '*château fort*' partagent aussi beaucoup de *valeurs* que peuvent assumer les propriétés d'objet. Ils se définissent comme des *constructions en pierre de grande taille* dont l'*investissement financier* est *coûteux*. Seuls leur *mode de construction* et la *fonction spécifique* diffèrent en vue de la *valeur*. L'aberration de *valeurs* s'explique par deux contextes culturels divergents dans lesquels s'inscrivent les deux objets - dans le cas présent, ils sont causés par différentes *circonstances historiques*. Le '*château fort*' a pour préoccupation principale la protection du souverain, donc la *défense* de la cour. Entre le Xe et XVe siècle, le '*château fort*' constitue l'échine d'un système de défense du territoire, la résidence *fortifiée* du seigneur.<sup>274</sup> Après la guerre de Cent Ans (1337 - 1453), qui cause des progrès de l'artillerie, le souci n'est plus à la défense, mais au confort. Le '*château*' a pour préoccupation principale maintenant la *représentation* du souverain. Le '*château*' devient une demeure aristocrate qui - maintenant *non-fortifiée* - peut s'ouvrir sur la campagne environnante.<sup>275</sup>

La représentation de la *Sémantique des cadres* montre en détail sur quel contexte formel  $K(G,M,W,I)$  et sur quelles incidences *objet-attribut-valeur* repose la proximité sémantique

<sup>273</sup> Cf. Mansour (1983), pp. 5/6.

<sup>274</sup> Informations supplémentaires tirées de l'article « château » de Wikipédia.fr.

<sup>275</sup> Informations supplémentaires tirées de l'article « château » et « guerre de cent ans » de Wikipédia.fr.

de deux objets ‘*château*’ et ‘*château fort*’, qui sont déterminés par un contexte culturel spécifique. La *Sémantique de cadres* permet aussi d’évaluer en détail sur quel *degré* (indiqué par le chiffre comparatif des propriétés partagées et non-partagées) repose la similarité sémantique. La méthode présente est ainsi plus fine et détaillée qu’une acceptation intuitive de similarité sémantique, qui repose souvent uniquement sur la considération des *propriétés partagées* par deux objets.

### Exemple 2 :

La *Sémantique des cadres* permet également d’étudier en détail la relation d’*analogie* entre deux objets qui figurent dans deux domaines sémantiques éloignées. Reprenons pour l’exemple la relation d’*analogie* entre ‘*avion*’ et ‘*oie*’, qui va nous permettre d’illustrer la fonction du cadre sémantique en tant que système conceptuel qui impose structure et cohérence sur certains aspects de l’expérience humaine.<sup>276</sup>

Les deux termes ‘*avion*’ et ‘*oie*’ apparaissent - comme nous l’avons déjà vu - dans deux entrées dictionnaires différentes (cf. chapitre 3.3.2). ‘*Avion*’ figure dans l’entrée 820 *Transports par air* et ‘*oie*’ figure dans l’entrée 297 *Oiseaux*. Ils dépendent de deux termes génériques différents, ‘*moyen de transport*’ et ‘*oiseau*’, qui inscrivent les termes dans une *relation genre-espèce*. Par contraste avec l’exemple précédent, l’héritage des propriétés génériques est de moindre importance pour l’étude de l’analogie, qui s’instaure entre deux *espèces* de classes différentes. Par conséquent, les cadres ‘*avion*’ et ‘*oie*’ sont reliés *implicitement* à leurs cadres génériques dans le schéma suivant (par la relation AKO)<sup>277</sup>. L’analyse se concentre surtout sur la *valeur* qui s’ajoute comme troisième élément au système d’attribution *incidence objet-attribut*. Le tableau 19 présente l’analyse du cadre sémantique pour ‘*avion*’ et ‘*oie*’.

---

<sup>276</sup> Cf. Konerding (1993), p. 21.

<sup>277</sup> Le signe AKO désigne la relation A-Kind-Of, en français « une espèce de ».

**Tableau 19 : Analyse du cadre sémantique de 'avion' et 'oie'** <sup>278</sup>

{	CADRE 4 : L'avion
	(m. avion)
	(AVION( <b>AKO</b> (MOYEN DE TRANSPORTS)))
	(ENVERGURE (80 m))
	(CHANGEMENT DE DIRECTION (par alignement des ailes))
	(VITESSE DE CROISIÈRE (600 km/h))
	(MASSE MAXIMALE AU DECOLLAGE (800 t))
{	(CONSOMMATION D'ENERGIE (1500 l/100 km))
	CADRE 5: L'oie
	(f. oie)
	(OIE ( <b>AKO</b> (OISEAU)))
	(ENVERGURE (160 cm))
	(CHANGEMENT DE DIRECTION (par alignement des ailes))
	(VITESSE DE CROISIÈRE (90 km/h))
	(MASSE MAXIMALE AU DECOLLAGE (7,5 kg))
	(CONSOMMATION D'ENERGIE (200 cal. /jour))

L'*analyse formelle des concepts* avait montré que 'avion' et 'oie' partagent une majorité de propriétés (cf. chapitre 3.3.2). La question qui se pose est de savoir quel impact a la *valeur* lors de l'établissement d'une analogie entre termes. Pour évaluer ceci, les propriétés éprouvées de l'*analyse formelle des concepts* ont été insérés dans le cadre conceptuel et ont été affectées de *valeurs numériques*.

Les deux termes 'avion' et 'oie' peuvent être alignés sur différentes propriétés partagées. La création d'une *source* (oiseau) et d'une *cible* (avion) peut, dans cette perspective, soutenir un raisonnement par analogie qui apporte des améliorations technologiques par l'apprentissage de la nature. Dans ce contexte, 'avion' et 'oie' peuvent être considérés comme analogues concernant leur 'envergure' qui peut atteindre une certaine longueur par rapport au corps. L'oie présente une envergure de 160/90 cm, l'avion une envergure de 80/70 m. <sup>279</sup> La comparaison des dimensions d'envergure pourrait optimiser le comportement en vol des avions futurs. 'Avion' et 'oie' peuvent être considérés comme

<sup>278</sup> **Tableau 19** de S. Reusch. Élaboré d'après un schéma proposé par Mansour (1983), p. 5 et d'après les informations trouvés dans le Thésaurus Larousse (1992), articles 297. Oiseaux (pp.194-197) et 820. Transports par air (pp.578/579).

<sup>279</sup> Informations exemplaires présenté dans cet exemple ont été tirés des articles Wikipédia « Wildgans », ainsi que « Airbus A380 », sur : [www.wikipedia.de](http://www.wikipedia.de).

analogues concernant leur *consommation d'énergie*. L'oie consomme 200 calories/jour en énergie, l'avion 1500 litres/100km. La comparaison de l'apport et de la consommation d'énergie pourrait optimiser les mesures d'économie d'énergies renouvelables que pourraient consommer futurs avions. 'Avion' et 'oie' peuvent être considérés comme analogues concernant leur *masse maximale de décollage* (7,5 kilos pour une oie, 500 tonnes pour un avion), leur *vitesse de croisière* (90 km/h pour une oie, 600 km/h pour un avion) et leur *changement de direction* (par l'alignement des ailes, à l'aide des muscles pour une oie ou à l'aide d'une commande conduite pour un avion).

La *Sémantique des cadres* montre que l'analogie entre une source 'oie' et une cible 'avion' livre une richesse d'idées pour pouvoir optimiser l'aviation des générations futures. Les *valeurs pragmatiques* jouent, dans cette perspective de raisonnement par analogie, le rôle du *fournisseur d'idées*. La comparaison d'*objets*, comme celle de 'oie' et 'avion', qui repose uniquement sur une comparaison d'*attributs*, comme l'existence d'une 'envergure', ne mène à rien pour la résolution d'idées. L'explication est simple, l'attribution d' 'ailes' aux '*objets volants*' a été opérée par l'homme il y a bien longtemps (p.ex. dans les travaux aéronautiques de Leonardo DaVinci). La comparaison d'*objets*, qui repose sur un système d'attribution élargi - celle de *l'incidence objet-attribut-valeur* - peut apporter des idées nouvelles. La comparaison d'objets, tel que 'oie' et 'avion', qui repose sur l'évaluation d'un attribut, tel que l' 'envergure', en combinaison avec une *valeur numérique*, tel que '160/90 cm' et '80/70m', peut lier à la résolution des problèmes techniques complexes de l'aéronautique. La description de concepts à l'aide d'un triple *objet-attribut-valeur* est une forme particulièrement raffinée de la représentation des connaissances au moyen d'un réseau sémantique, qui est le fondement d'un thésaurus. Elle permet une '*analyse formelle de concepts*' en rapport avec la réalité de l'utilisateur du thésaurus.

### Exemple 3 :

Dans sa publication « *A structural approach to analogy* » (1983) Hormoz Mansour propose une seconde méthode d'analyse pour l'analogie. Il spécifie la méthode de Charles J. Fillmore et se concentre sur l'idée que deux *cadres* peuvent se relier par un seul *élément cadre*.<sup>280</sup> Pour une analyse de la relation d'*analogie* on réduit le '*contenu informationnel*'

---

<sup>280</sup> Cf. Mansour (1983), pp.5/6.

des cadres à la description d'un seul attribut d'objet, sur lequel repose l'association de termes. Dans cette perspective, un seul 'goal' (lien de signification) permet d'évaluer différents 'slots' (types d'attribut) possibles d'une même propriété.<sup>281</sup> Par conséquent, la relation d'analogie entre deux objets, comme 'avion' et 'oie', peut être décrit grâce à un seul *critère d'analyse* en vue d'une *propriété spécifique partagée* :

(simple-analog « Avion » « Oie »)

\*\*\*GOAL ?

SLOT : CHANGEMENT DE DIRECTION

\*\*\* CHANGEMENT DE DIRECTION of AVION is caused by ALIGNEMENT DES AILES and CHANGEMENT DE DIRECTION of OIE is caused by ALIGNEMENT DES AILES - therefore analogy of CHANGEMENT DE DIRECTION between AVION and OIE.<sup>282</sup>

La seconde méthode d'analyse repose sur un cas spécifique d'alignement de termes, que nous avons défini dans ce mémoire, alternativement, comme case de départ pour la *formation de métaphores* ou comme case de départ pour le *raisonnement par analogie* (cf. chapitre 3.2.3). Hormoz Mansour se concentre dans son approche sur le *raisonnement par analogie*, qu'il situe dans un *contexte formel* de domaines sémantiques pour pouvoir effectuer une *analyse structurelle* du phénomène. Dans la citation suivante il expose son point de vue sur cette question :

« There are multiple sorts of reasoning by analogy between two domains; the one with which we are concerned is a type of *contextual analogy*. (...) A domain is contextually analogous to another, if at one precise moment in the analysis, there exists similarity between it and another domain which is the same state and context. In this way, we can observe that at one level of analysis [*slot*] an analogy exists between two domains [*frames*]. (...) Therefore, we have a concept of analogy which is essentially based on *structure*. This means when we study this sort of reasoning, we must be aware of both the *level* [*slot*] and the *context* [*frame*]. »<sup>283</sup>

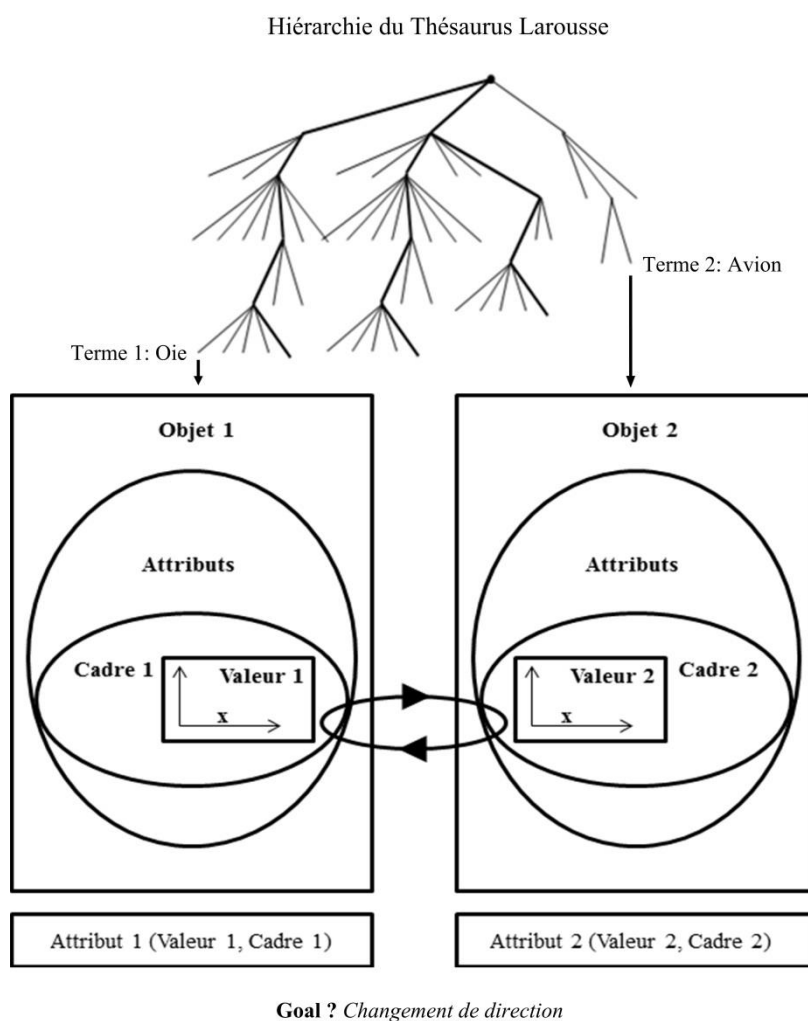
---

<sup>281</sup> Cf. Mansour (1983), pp. 5-10.

<sup>282</sup> Cf. Mansour (1983), Schéma, p. 6.

<sup>283</sup> Cité d'après Mansour (1983), p. 1 (première phrase) et p. 3 (reste), ajouts entre guillemets et mises en italique fait par S. Reusch.

Le schéma suivant montre - pour le Thésaurus Larousse - ce que Mansour appelle une *approche structurale de l'analogie*.<sup>284</sup> Dans un thésaurus, la *relation d'association entre termes* établit un lien entre deux *termes* qui figurent dans deux domaines sémantiques différents (cf. chapitre 1.3.3).<sup>285</sup> Elle indique un *lien de signification* (inform. 'goal') par lequel on associe mentalement deux *objets* qui figurent dans différents *contextes*.<sup>286</sup> Le *lien de signification* se construit - dans le cas de l'analyse structurale - sur une *incidence attribut-valeur* se produit dans deux *cadres conceptuels* en alignement.



**Figure 15: Le raisonnement par analogie**<sup>287</sup>

Le schéma présente l'analogie comme une forme de comparaison entre représentations qui dépend des *similitudes de structure* (exprimé par une *incidence objet-attribut-valeur*) qui

<sup>284</sup> Tableau de S. Reusch, informations tirés de Mansour (1983), p. 3 et du Thésaurus Larousse (1992), article 297.7 Oiseaux (pp.194-197) et 820.1 Transports par air (pp.578/579).

<sup>285</sup> Site du Politique d'édition TAG, consultable sur :[http://www.thesaurus.gouv.qc.ca/tag/acc\\_poltq\\_editn.do](http://www.thesaurus.gouv.qc.ca/tag/acc_poltq_editn.do)

<sup>286</sup> Cf. Bianucci (1992), p.78.

<sup>287</sup> Figure 15 de S.Reusch. Élaboré sur un schéma proposé par Mansour (1962), p. 3 qui a été littéralement adopté et élargi par les informations trouvés dans le Thésaurus Larousse (1992), Sommaire, pp. xv-xxi.

sont hiérarchisées par l'importance dans la réalisation d'un but (exprimé par un '*goal*').<sup>288</sup> Un contexte est présenté comme composée d'objets, ayant des attributs et des valeurs précis. Le formalisme de représentation dépend du *type prédicat argument*, qui s'exprime ainsi : « Une *oie* change de direction *par l'alignement des ailes*. ».<sup>289</sup>

L'approche structurale de Hormoz Mansour, nous fait comprendre l'analogie comme un '*transfert d'apprentissage*' et comme une '*heuristique de découverte*'<sup>290</sup>, pour lesquels une '*mise en correspondance*' de deux objets constitue l'étape principale et caractéristique.<sup>291</sup> Dans cette étape, la détermination d'une '*source*' (*oie*) et d'une '*cible*' (*avion*) fait de l'analogie un mécanisme puissant pour la cognition humaine.<sup>292</sup> Marvin L. Minsky, l'un des grands ancêtres de la *Sémantique des cadres* avec Charles J. Fillmore, considère la métaphore, et la forme du raisonnement y attachée, comme une ressource inéluctable pour la solution des problèmes. Pour lui, l'analogie est plus qu'un processus mental d'*alignement de termes*. L'analogie un mécanisme qui rend possible de comprendre une *idée* (ou un *objet*) en termes d'une autre *idée* (ou d'un autre *objet*).<sup>293</sup> Dans sa publication « Jokes and their Relation to the Cognitive Unconscious » (1980) Minsky explique l'importance des *systèmes de cadres* (angl. *frame-systems*) pour l'apprentissage par analogie:

« All intelligent persons also possess some larger-scale frame-systems whose members seemed at first impossibly different - like *water* with *electricity*, or *poetry* with *music*. Yet many such analogies - along with the knowledge of how to apply them - are among our most powerful tools of thought. They explain our ability sometimes to see one thing - or idea - as though it were another, and thus to apply knowledge and experience gathered in one domain to solve problems in another. It is thus that we transfer knowledge via the paradigms of Science. We learn to see *gases* and *fluids* as *particles*, *particles* as *waves*, and *waves* as *envelopes of growing spheres*. »<sup>294</sup>

Le Thésaurus Larousse est - par sa structure conceptuelle - en mesure de soutenir un tel processus d'apprentissage. La conception de l'ouvrage repose sur des *systèmes conceptuels* ce qui rend possible l'*association d'idées* par mesure de l'*association de termes*.

<sup>288</sup> Cf. Sander (2000), p. 21.

<sup>289</sup> Cf. Sander (2000), p. 10.

<sup>290</sup> Terminologie utilisée par Sander (2000), Préface, p. iii.

<sup>291</sup> Terminologie utilisée par Sander (2000), p. 16.

<sup>292</sup> Cf. Sander (2000), Introduction, p. iii.

<sup>293</sup> Cf. Holyoak / Thagard (1995), p.5; cf. Fillmore (1980), p. 9.

<sup>294</sup> Cité d'après Minsky (1980), p.9. Mise en italique par S.Reusch.

### 3.4.2 Le cadre - représentation des connaissances du monde

La théorie de la *Sémantique des cadres* ne propose pas uniquement une structure pour la comparaison de deux concepts - comme l'*analyse formelle de concepts*. Au niveau des mots, elle utilise le *cadre* comme outil de sémantique lexicale, qui relie la *description sémantique* des termes aux *connaissances du monde* de l'utilisateur, en fournissant des informations supplémentaires (*savoir contextuel*) sur des individus, objets ou situations dénotés par le terme.<sup>295</sup> Dans la publication « Frames in Linguistic Descriptions » (1985) Haldur Oim et Madis Saluveer expliquent la fonction du *cadre* - sur le seuil entre classification lexicale et réalité linguistique - en termes suivants :

« Frames were primarily seen as representational tools by means of which purely linguistic knowledge could be connected with relevant non-linguistic commonsense knowledge. On the more concrete level, the use of frames in linguistics is primarily associated with the semantic description of words, of lexicon. So the bulk of theoretical discussion and concrete examples concerning the use of frames in linguistics is - explicitly or implicitly - about frames as conceptual tools of lexical semantics, frames as means of including into the semantic descriptions of words also relevant types of common sense knowledge about objects or situations referred to by a corresponding word [...]. »<sup>296</sup>

Cette définition nous rapproche d'un modèle encyclopédique du lexique, qui nous montre de plus en plus clairement qu'un *cadre* doit être compris, au sein d'un thésaurus, comme une description - déterminée par des normes culturelles ou institutionnelles - de la connaissance du monde, attachée au savoir sur des objets référentiels.<sup>297</sup> Comment se constitue un tel *cadre* en vue d'un objet déterminée par les normes spécifiques ? Prenons l'exemple du terme '*rectangle*', qui figure dans l'article 146 *Géométrie* du Thésaurus Larousse.<sup>298</sup> Ce terme est déterminé par une partie des mathématiques, qui définit la nature de différentes figures géométriques en vue d'un espace dans lequel elles apparaissent.<sup>299</sup> Par conséquent, on ne peut définir le terme '*rectangle*' qu'après avoir compris la notion d'*espace* et de *figure* (le *savoir contextuel*) en géométrie.<sup>300</sup> Le schéma suivant montre la définition géométrique d'un rectangle :

---

<sup>295</sup> Cf. Taylor (1998), pp. 94/95.

<sup>296</sup> Cité d'après Oim/Saluveer (1985), pp. 296/297.

<sup>297</sup> Cf. Konerding (1993), p. 47.

<sup>298</sup> Exemple tiré du Thésaurus Larousse (1992), article 146.5 Géométrie, p. 79.

<sup>299</sup> Cf. Définition de « géométrie » tirée du Dictionnaire du Centre des Ressources Textuelles et Lexicales, consultable sur : <http://www.cnrtl.fr/definition/géométrie>.

<sup>300</sup> Cf. Monneret (2004), pp. 30-35 ; cf. Reusch (1988), p. 103.



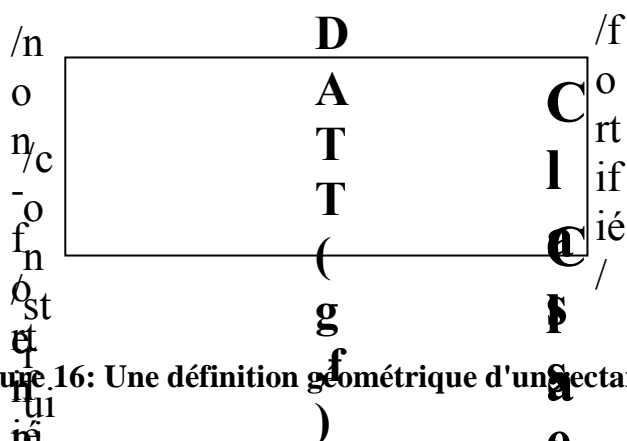


Figure 16: Une définition géométrique d'un rectangle <sup>301</sup>

Ce schéma peut être représenté par le *cadre* suivant <sup>302</sup> :

1. (Type de cadre : Rectangle)
2. (Numéro de cadre : 1)
3. (Côtés :  $a = 6$  ;  $b = 3$  ;  $c = 6$  ;  $d = 3$ )
4. (Sommets : A, B, C, D)
5. (Angles :  $\alpha = 90$  ;  $\beta = 90$  ;  $\gamma = 90$  ;  $\delta = 90$ )
6. (Equations:  $\alpha + \beta + \gamma + \delta = 360$ )

Le *cadre conceptuel* du terme 'rectangle' se crée par rapport à d'autres figures géométriques existantes. Ainsi, les deux premiers *slots* sont utilisés pour l'identification du *cadre*, par type et par numéro. <sup>303</sup> Le *cadre* présente les caractéristiques que peuvent montrer différentes figures géométriques, par divers '*slots*'. Des figures telles que le 'pentagone' ou l' 'octogone', sont définies en mathématiques par une différente longueur de côtés (*slot 3*), un différent nombre de sommets (*slot 4*), une différente valeur d'angles (*slot 5*) et des équations mathématiques différentes (*slot 6*). Le *cadre* du terme 'rectangle' contient ainsi toutes les informations essentielles pour la représentation d'un objet géométrique. <sup>304</sup> La définition du concept 'rectangle' s'opère par l'attribution de différentes *valeurs* de propriétés- comme l'angle droit - qui sont communément admises en domaine de géométrie. Le modèle encyclopédique du lexique adopté ici, définit le *cadre*

<sup>301</sup> Figure 16 cité d'après Reusch (1988), p. 102.

<sup>302</sup> Cadre cité d'après Reusch (1988), p. 102.

<sup>303</sup> Cf. Minsky (1975), pp. 212/213.

<sup>304</sup> Cf. Reusch (1988), p. 103.

comme une *description standardisée des connaissances du monde*.<sup>305</sup> Fillmore élargit l'idée dans sa publication « Frames and the Semantics of Understanding » (1985), en disant qu'un *cadre* présenterait « *particular structured understandings of cultural institutions, beliefs about the world, shared experiences, standard ways of doing things and ways of seeing things.* »<sup>306</sup>. Les membres d'une communauté de culture partagent tout - des *normes* jusqu'aux *croyances*, des *décrets d'application préfixés* jusqu'aux *expériences individuelles* - à base de *cadres conceptuels*, qui mettent de l'ordre dans le flux d'informations. Pour la classification opérée dans un thésaurus, cela signifie : Un *cadre* qui montre des descriptions totalement fixées pour un concept - comme celui pour le terme '*rectangle*' - ne peut pas être attendu pour des objets ou situations de nature plus complexe.<sup>307</sup>

Si l'on prend en considération la déclaration suivante : « *A frame is any system of concepts related in such a way that to understand any one concept it is necessary to understand the entire system.* »<sup>308</sup>, la question qui se pose est celle de savoir quel impact a le *système de cadres* pour l'organisation conceptuelle d'un thésaurus.

### 3.4.3 Le Thésaurus - une conception dépendante du système de cadres

Dans un thésaurus, l'utilisateur serait sûrement surpris de trouver le mot *printemps* introduit dans un premier article et les mots *été*, *automne* et *hiver* distribués au hasard dans trois autres articles.<sup>309</sup> Il ne s'attendrait pas non plus à trouver les mots *mère*, *père*, *fille*, *fil*, séparés les uns des autres, dans des articles distincts. Ces mots forment des groupes que l'utilisateur doit pouvoir consulter ensemble.<sup>310</sup> Dans chacun de ces exemples, les mots sont des *représentants* d'une certaine schématisation lexicale, cohérente et unique, de connaissance ou de l'expérience humaine. « *A frame is any system of concepts related in such a way that to understand any one concept it is necessary to understand the entire system.* »<sup>311</sup> Comprendre le sens du mot '*printemps*' signifie donc, en quelque sorte, comprendre le sens de tous les autres mots '*été*', '*automne*' et '*hiver*', à base d'un système

<sup>305</sup> Cf. Konderding (1993), p. 47.

<sup>306</sup> Cité d'après Fillmore (1985), p. 231.

<sup>307</sup> Cf. Konderding (1993), p. 47.

<sup>308</sup> Cité d'après Petruck (1996), p.1.

<sup>309</sup> Cf. Fillmore/Atkins (1992), pp. 77/78 (exemples pour *day names*).

<sup>310</sup> Cf. Fillmore (1985), p. 223.

<sup>311</sup> Cité d'après Petruck (1996), p.1.

cohérent.<sup>312</sup> Il semble donc naturel de s'attendre à ce que les concepteurs d'un thésaurus regroupent les mots ensemble, dans un article de dictionnaire, intitulé '*saisons*' ou '*famille*'.<sup>313</sup>

Ce qui tient ces groupes de mots ensemble est, selon Charles J. Fillmore, le fait d'être fondée sur une '*schématisation de l'expérience*' et le fait d'être motivée par une '*standardisation des connaissances*' - et c'est pour rendre compte de ces faits que peut être construit un *cadre* comme représentation sémantique.<sup>314</sup> Si nous voulons exprimer, par exemple, notre compréhension de '*saisons*' ou d'autres termes reliés, nous pouvons faire appel à un seul *cadre*. Il se compose d'une compréhension (1) du cycle naturel créé par l'orbite annuelle du soleil, (2) des moyens normés pour reconnaître qu'une saison se termine et une prochaine saison commence, (3) du cycle civil, de l'année calendaire (4) de la pratique de notre culture d'assigner à différentes étapes du cycle annuel des fêtes spécifiques [La fête du premier mai ; la fête de la mi-été (danois : *Midsommer*) ; la fête des récoltes et la fête de Noël ; etc.].<sup>315</sup> La conscience implicite de cette organisation du monde - qu'elle soit physique ou sociale - permet d'organiser le riche matériel lexical dans des *articles*, plus précisément dans des *paragraphes reliés* au sein du thésaurus. Ils forment un *cadre sémantique*, qui englobe des noms communs, tel que '*saison*', '*mois*' ou '*semaine*', des dérivés adjectivaux, tel que '*estival*' ou '*hibernal*', les noms des mois, tel que '*janvier*', '*février*', '*mars*', etc., et des termes spéciaux tel que '*vacances d'été*', '*époque de Noël*' ou '*avent*'.<sup>316</sup> Un tel *cadre* représente l'organisation particulière des connaissances - une condition préalable à notre capacité de comprendre la nature des *termes* et des *relations sémantiques* dans un thésaurus. (cf. chapitre 1.3.2 et 1.3.3).<sup>317</sup>

---

<sup>312</sup> Cf. Fillmore (1985), p. 223.

<sup>313</sup> Cf. Thésaurus Larousse (1992), articles 187. Saisons (p. 103) et 678. Famille (p. 466).

<sup>314</sup> Cf. Fillmore (1985), p. 223.

<sup>315</sup> Cf. Fillmore (1985), pp. 223/224, le cadre interprétatif de *week-days*.

<sup>316</sup> Cf. Fillmore (1985), pp. 223/224, le cadre interprétatif de *week-days*.

<sup>317</sup> Cf. Fillmore (1985), p. 224.

## 3.5 Mesure de la similarité conceptuelle dans un thésaurus

### 3.5.1 Le Thésaurus : Un instrument fiable pour la mesure de similarité

La compréhension de la *similarité sémantique*, c'est-à-dire la détection de la liaison entre concepts, est une capacité humaine. Il est évident, pour l'homme, que les mots *marteau* et *tournevis* sont liés, beaucoup plus que les mots *tournevis* et *voiture* de toute façon.<sup>318</sup> Mais il est difficile pour la linguistique de formaliser ce lien conceptuel, car rien en surface, ne permet de le décider.<sup>319</sup> Pour capturer ces '*distances intuitives*' ressenties pour les paires *tournevis* - *marteau* et *tournevis* - *voiture*, il faut recourir à une ressource lexicale qui présente des connaissances formalisées dans une structure hiérarchique.<sup>320</sup>

Un thésaurus est une ressource lexicale qui repose sur des connaissances formalisées. Il capture le consensus de plusieurs rédacteurs en vue de l'organisation conceptuelle d'un *domaine de connaissances*.<sup>321</sup> L'ouvrage se compose d'une multitude de *domaines* (ou *champs conceptuels*) qui sont intégrés par des *concepts* (ou *termes génériques*) dans une structure hiérarchique. La structure s'élabore - au fur et à mesure - que l'on classe les concepts dans une hiérarchie. Ce procédé nécessite une évaluation de chaque *terme générique* (concept) en fonction des *termes spécifiques* (mots et locutions) qu'il subsume. 'Évaluer' signifie, dans cette perspective, que chaque concept est classé en fonction de la *notion* qu'il dénote en vue d'un *domaine de connaissances*. La structure d'un thésaurus est donc organisée par une « *hiérarchie des notions* » (cf. chapitre 1.3.4).

La « *hiérarchie des notions* » est une structure qui supporte naturellement l'évaluation de similarité sémantique. Dans cette structure, les *termes spécifiques* sont regroupés selon des caractéristiques définitionnelles qu'ils partagent [*être un outil* et *être un moyen de transport*] en *notions conceptuelles* qui régissent les articles de dictionnaire [799. *Outil* et 817. *Automobile*].<sup>322</sup> La hiérarchie des notions est un *système* qui permet de visualiser les *liens hiérarchiques* qui s'établissent au sein du *réseau sémantique* d'un thésaurus (cf. figure 17).

---

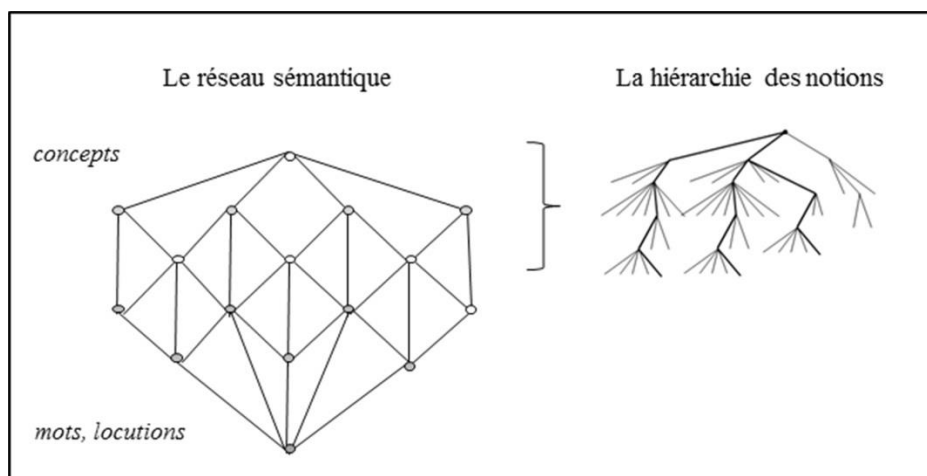
<sup>318</sup> Cf. Ventresque (2006), p. 372 ; cf. Jarmasz /Szpakowicz (2003), p. 212.

<sup>319</sup> Cf. Ventresque (2006), p. 372.

<sup>320</sup> Cf. Gandon et al. (2008), p.87 ; cf. Jarmasz /Szpakowicz (2003), p. 212.

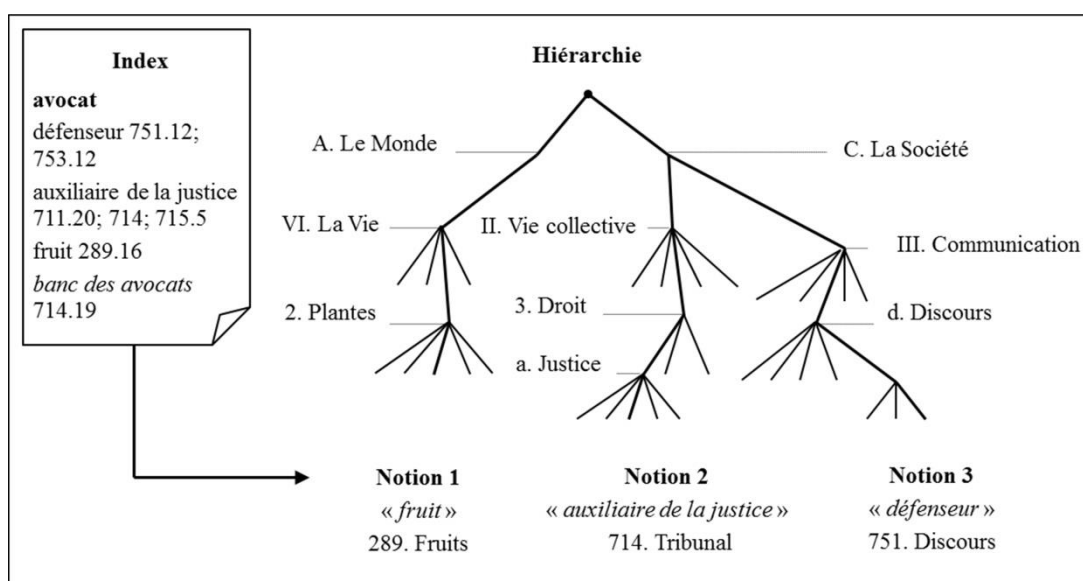
<sup>321</sup> Cf. Jarmasz /Szpakowicz (2003), p. 213.

<sup>322</sup> Cf. Gandon et al. (2008), p.87. Informations tirés du Thésaurus Larousse (1992), articles 799. Outil (p. 555) et 817. Automobile (p.574).



**Figure 17: La hiérarchie des notions, un système qui reflète le réseau sémantique**<sup>323</sup>

Dans ce système chaque terme appartient à une catégorie conceptuelle qui fixe sa *notion*. Cette notion définit la nature du terme au sein de son voisinage sémantique.<sup>324</sup> Prenons l'exemple de l'homonyme '*avocat*' pour expliquer le système. Le terme '*avocat*' appartient, dans le Thésaurus Larousse, à trois catégories qui définissent sa *notion conceptuelle* de manière différente (cf. figure 18).



**Figure 18: Classement conceptuel du terme polysémique « avocat » dans le T.L.**<sup>325</sup>

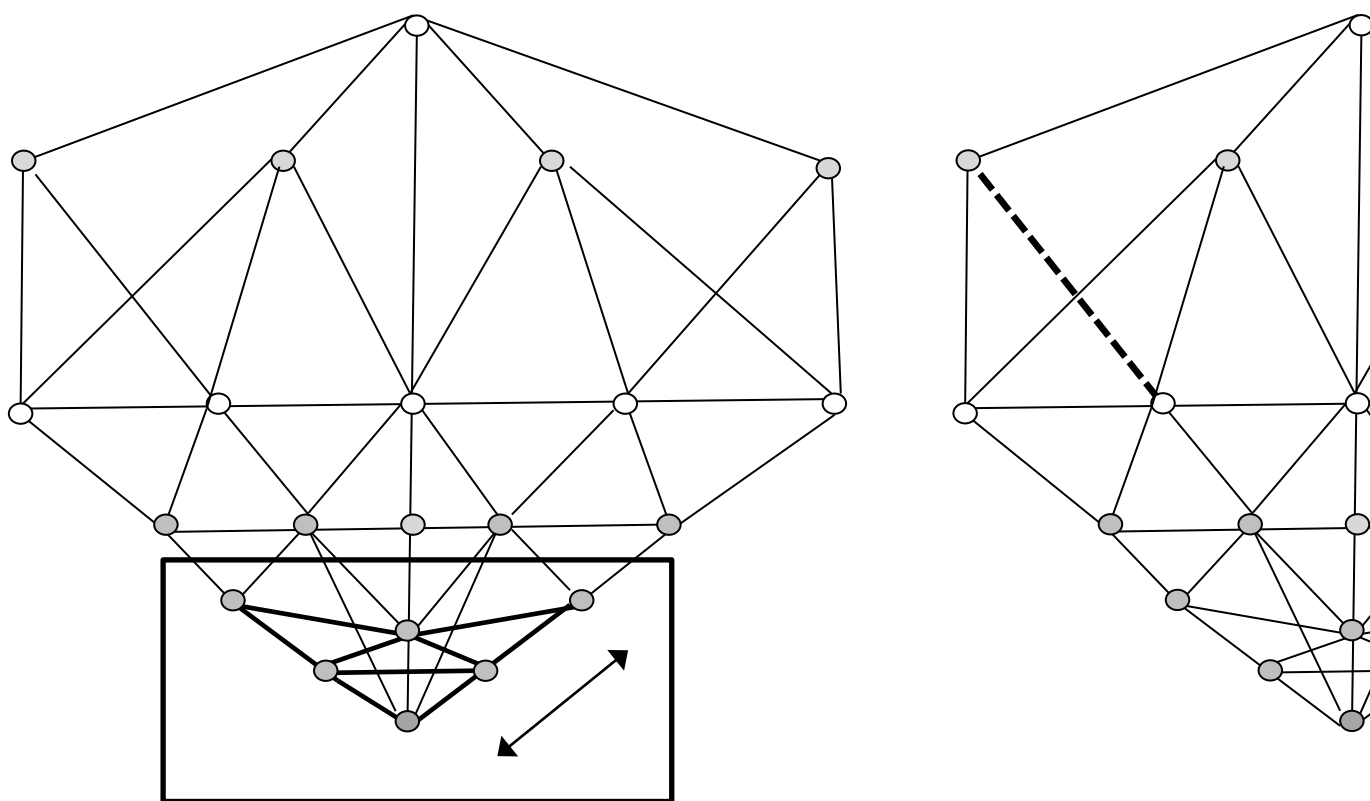
La catégorie '*plantes*' qualifie d'*avocat* un *fruit exotique* qui figure parmi d'autres, comme l'ananas, la banane ou la figue.<sup>326</sup> La catégorie '*justice*' qualifie d'*avocat* un *auxiliaire de*

<sup>323</sup> Figure 17 de S.Reusch. Élaboré à l'aide des informations préalablement utilisés dans figure 8.

<sup>324</sup> 1.2.2 Constitution d'un thésaurus : [http://www.sigb.net/doc/fr\\_FR/html-user/ch17.html](http://www.sigb.net/doc/fr_FR/html-user/ch17.html)

<sup>325</sup> Figure 18 de S.Reusch. Élaboré d'après les informations trouvés dans l'index pour l'item « avocat » (p. 671) et dans les articles 289. Fruits (p. 186), 714. Tribunal (p. 488) et 751. Discours (p. 518) du Thésaurus Larousse (1992).

*la justice* qui figure au tribunal parmi le juge, le procureur général ou le témoin.<sup>327</sup> La catégorie '*discours*' qualifie d'avocat un *défenseur de la bonne cause* comme l'orateur, le prédicateur ou l'avocat.<sup>328</sup> Les exemples montrent : Qu'un thésaurus permet de situer un terme - tel que '*avocat*' - dans plusieurs contextes.<sup>329</sup> À l'intérieur de chaque *contexte* un terme se définit de manière *univoque* par rapport à d'autres termes de sens proche, par rapport à *ananas* (1), *juge* (2) ou *orateur* (3). Dans le thésaurus, une proximité dans le *sens* peut ainsi se traduire par une proximité dans l'*espace* (cf. figure 19).



**Figure 19: Rapprochement des termes dans l'espace sémantique d'un thésaurus**<sup>330</sup>

Le système de catégorisation classe souvent deux termes sémantiquement proches dans un article de dictionnaire, dans le schéma symbolisé par une *case* (cf. figure 19-A). Néanmoins, il permet aussi de rapprocher deux termes sémantiquement proches qui appartiennent à deux articles différents, par un renvoi de terme à terme - dans le schéma symbolisé par une *ligne pointillée* (cf. figure 19-B).

<sup>326</sup> Cf. Thésaurus Larousse (1992), article 286. Fruits, paragraphe 16, p. 186.

<sup>327</sup> Cf. Thésaurus Larousse (1992), article 714. Tribunal, paragraphes 9 à 15, p. 488.

<sup>328</sup> Cf. Thésaurus Larousse (1992), article 751. Discours, paragraphe 12, p. 518.

<sup>329</sup> Cf. Ventresque (2006), p. 372.

<sup>330</sup> Figure 19 de S.Reusch. Élaboré d'après les expériences faits lors de la consultation du Thésaurus Larousse (1992) à la recherche des termes spécifiques et des termes associés d'un concept, comme « maison », « avion », etc.

Le schéma met en évidence qu'un thésaurus n'est pas *formellement* défini sur une hiérarchie des notions (cf. figure 17). La hiérarchie des notions est un *consensus* d'une rédaction - comme celle existante dans la Maison d'édition Larousse - en vue d'une organisation adoptée dans l'établissement d'un corpus de dictionnaire. La hiérarchie reflète alors l'organisation d'un ouvrage unique. Une organisation qui repose sur une vision partagée par plusieurs experts, formés dans la perception et le classement des phénomènes linguistiques (cf. chapitre 3.1.2).<sup>331</sup> Le *contexte formel* ainsi établi utilise pour la classification des *concepts*, qui considèrent les *objets* selon une perspective spécifique. Les concepts leur assignent *explicitement* uniquement les propriétés et les relations qui reflètent l'organisation lexicale adoptée par les experts (cf. tableau 20).<sup>332</sup>

**Tableau 20 : Un renvoi entre différentes notions d'avocat, le reflet d'un consensus adopté**<sup>333</sup>

**753. Rhétorique**

12 Orateur, rhéteur [litt.] ; *avocat* 714, parlementaire, tribun ; prédicateur. - Speaker (fem. : speakerine). - Écrivain 762, poète ; rhétoriqueur [vx.] ; félibre [région.]. - Chambre de rhétorique [hist.].



**714. Tribunal**

13 *Avocat de la défense*, défenseur. - avocat commis d'office, avocat plaidant, avoué plaidant. - Avocaillon ; arg. et péj. : babillard, bavard ; baveux, bêcheur, blanchisseur. - Avocat de la Cour, avocat au Conseil d'état, avocat à la Cour de cassation. - Maître ou, abrég. M<sup>e</sup> (*Maître Untel, avocat à la Cour*).

Le réseau sémantique est le *fondement* de chaque thésaurus. Il organise *implicitement* les mots et les locutions (appelés *termes*), ainsi que leurs concepts attachés, en '*domaines sémantiques*' (cf. figure 17 et 18). Le *contexte formel* ainsi complété permet de caractériser chaque *objet* par-dessus les catégories d'expériences prédéfinies.<sup>334</sup> Le réseau sémantique permet d'associer des termes qui appartiennent à des domaines sémantiques différents par diverses relations : hyperonymie/hyponymie, antonymie, synonymie ou analogie, qui attribuent aux *objets* - par adoption de diverses perspectives - des *propriétés* variés. L'organisation du réseau sémantique situe les termes [indiquant les *objets*] et les relations entre termes [indiquant les *propriétés d'objet*] au même plan. Ceci permet à l'utilisateur du

<sup>331</sup> Entretien avec Daniel Péchoin, le 20. Octobre 2010.

<sup>332</sup> Cf. Wille (2005), p. 12.

<sup>333</sup> **Tableau 20** de S. Reusch. Élaboré d'après les informations trouvés dans le Thésaurus Larousse (1992), articles 753. Rhétorique (p.519) et 714. Tribunal (p.488).

<sup>334</sup> Cf. Wille (2005), p. 12.

thésaurus de discerner la *nature complexe* d'un concept [constituée en *extension* et en *intension*] (cf. chapitre 3.1.1).<sup>335</sup> Le réseau sémantique permet aussi de rapprocher plusieurs *concepts*, dans une perspective sémantique. L'*association de termes* montre qu'un concept ne saisit jamais *directement* la réalité. Il invite à être incorporé, examiné et assimilé au schéma cognitif de l'utilisateur - un schéma formé par ses propres expériences linguistiques, culturelles et anthropologiques.<sup>336</sup>

Un thésaurus présente une *organisation consensuelle* du lexique de la langue naturelle, qui permet de comprendre qu'un concept est un élément de notre compréhension du monde. Un élément qui contient des *hypothèses implicites* et *explicites* sur les objets (ou événements), leur causes, leurs caractéristiques et leurs relations.<sup>337</sup> Il organise les concepts dans un *contexte formel*, dans un réseau hiérarchique avec une multitude d'autres concepts. Et, il les intègre dans une *sous-structure*, un réseau sémantique interne qui instaure, dans une *perspective contextuelle*, des relations multiples entre les concepts formels.<sup>338</sup> Cette organisation fait d'un thésaurus un instrument fiable pour l'identification et la mesure de la *similarité sémantique* à base de l'organisation conceptuelle du corpus.

L'identification de la similarité sémantique est un champ d'investigation fortement recommandé dans le domaine de l'informatique (notamment du *Web sémantique*) et dans le domaine de la recherche de l'information.<sup>339</sup> Depuis les années 1980, plusieurs mesures – à base des ontologies – ont été proposées dans ce sens.<sup>340</sup> Dans le chapitre présent, nous nous intéresserons à l'application des *mesures de similarité* sur les termes issus du Thésaurus Larousse. Un premier objectif sera de détecter le *degré de similitude* entre termes et la *nature du lien sémantique* qui s'instaure. Un deuxième objectif sera d'évaluer le *degré de similitude* en vue de la *position des termes* dans le corpus. Par les deux approches, nous serons en mesure de capturer cette affinité essentielle entre la *forme* et le *contenu*, qui constitue la nature du dictionnaire.

### 3.5.2 Deux approches de mesures de similarité

La similarité entre termes d'un thésaurus a été un champ d'investigation pour de nombreux chercheurs. Deux grandes approches peuvent être distinguées : Une approche basée sur la

<sup>335</sup> Cf. Sémanticlopédie, Article : « Réseaux sémantiques » (de Patrick Saint-Dizier), consultable sur internet.

<sup>336</sup> Cf. Wille (2005), p. 12.

<sup>337</sup> Cf. Wille (2005), p. 7.

<sup>338</sup> Cf. Wille (2005), pp. 7/8.

<sup>339</sup> Cf. Khatraoui et al. (2008), p.121 ; cf. Slimani et al. (2007), p.1 ; cf. Harrathi / Calabretto (2010), p. 1.

<sup>340</sup> Cf. Slimani et al. (2007), p.1 ; cf. Ventresque (2006), p. 372 ; cf. Harrathi / Calabretto (2010), p. 1.



*distance sémantique* [Rada et al. 1989] et une approche basée sur la *proximité sémantique* [Wu/Palmer 1994, Resnik 1995].<sup>341</sup> La première mesure prend en considération la *structure hiérarchique du thésaurus*, la seconde prend le *contenu informationnel* des concepts, pour calculer la similarité sémantique.<sup>342</sup>

### 3.5.2.1 Approche basée sur la distance sémantique

La *mesure basée sur la distance sémantique* considère le thésaurus comme un *arbre* (au sens de la *théorie des graphes*) qui présente des *nœuds* (concepts) et des *arêtes* (chemins entre concepts).<sup>343</sup> La similarité est définie par rapport à cet arbre. Elle est définie par rapport à la *longueur des chemins* (calculée en *nombre d'arêtes*) qui séparent deux concepts dans la hiérarchie.<sup>344</sup> La méthode, telle qu'elle est considérée par les experts comme Rada, part d'une « intuition naturelle »<sup>345</sup> : plus deux concepts sont éloignés *physiquement*, au sein d'une hiérarchie sémantique, plus ils sont éloignés *sémantiquement*.<sup>346</sup> La hiérarchie sémantique sert de référence au calcul de distance. C'est par rapport à elle que se définit la notion même de *distance sémantique*.

#### Définition de la distance sémantique :<sup>347</sup>

Une distance  $d$  est dite *sémantique* si elle vérifie les propriétés suivantes :

1) si un concept  $C_1$  est identique à un concept  $C_2$  alors la valeur de distance entre les deux concepts est nulle (cas de distance minimale ou de synonymie lexicale)<sup>348</sup> :

$$\text{si } C_1 \equiv C_2 \text{ alors } d(C_1, C_2) = 0$$

2) si un concept  $C_1$  diffère complètement d'un concept  $C_2$  alors la valeur de distance entre les deux concepts est seize (cas de distance maximale)<sup>349</sup> :

$$\text{si } C_1 \neq C_2 \text{ alors } d(C_1, C_2) = 16$$

<sup>341</sup> Cf. Rada et al. (1989), pp. 17-30 ; cf. Wu/Palmer (1994), pp. 133-138 ; cf. Resnik (1995), pp. 448-453 ; cf. Resnik (1999), pp. 95-130 ; cf. Harrathi / Calabretto (2010), p.5 ; cf. Khatraoui et al. (2008), pp.122/123.

<sup>342</sup> Cf. Harrathi / Calabretto (2010), p.1.

<sup>343</sup> Cf. Cogis/Robert (2003), p.7 ; cf. Khatraoui et al. (2008), p.122 ; cf. Slimani et al. (2007), p. 3 ; cf. Ventresque et al. (2008), p.33.

<sup>344</sup> Cf. Harrathi / Calabretto (2010), p.6 ; cf. Khatraoui et al. (2008), p.122.

<sup>345</sup> Cité d'après Mazuel (2008), p.28 ; cf. aussi Ventresque et al. (2008), p.33.

<sup>346</sup> Cf. Mazuel (2008), p.28.

<sup>347</sup> Cf. Maala et al. (2007), p.10.

<sup>348</sup> Formule citée et texte adapté d'après Maala et al (2007), p.10.

<sup>349</sup> Formule de S. Reusch.

La distance  $d$  est dit *sémantique* car elle permet d'illustrer certaines relations sémantiques intuitives, relatives à la ressemblance entre les concepts. En même temps, elle représente une *distance mathématique* qui doit vérifier deux axiomes<sup>350</sup> :

1) La valeur de distance 0 est attribué uniquement à deux termes représentant exactement le même concept. Les termes sont désignés comme *synonymes*, alors que la valeur conceptuelle partagée est désignée comme *identique*<sup>351</sup> :

$$d(C_1, C_2) = 0 \Leftrightarrow C_1 = C_2$$

2) Le calcul de distance repose sur une symétrie entre concepts, c'est-à-dire que la *valeur de distance* reste identique si on change de direction. Le chemin dans la direction  $C_1 \rightarrow C_2$  est de la même longueur que le chemin parcouru dans la direction  $C_2 \rightarrow C_1$ <sup>352</sup> :

$$d(C_1, C_2) = d(C_2, C_1)$$

La mesure du *edge counting* proposée par Rada et al. [1989] utilise une métrique  $dist(C_1; C_2)$ , qui mesure le nombre d'arêtes séparant les deux concepts  $C_1$  et  $C_2$  par le chemin le plus court dans la hiérarchie. Plus deux concepts sont éloignés, moins ils sont similaires.<sup>353</sup> La mesure est donnée dans la formule suivante<sup>354</sup> :

$$Sim_{rada}(C_1; C_2) = \frac{1}{1 + dist(C_1; C_2)}$$

La mesure de distance sémantique nécessite une *structure* pour opérer les calculs. La *hiérarchie des notions* dans un thésaurus fournit une base adéquate. Elle permet d'explicitier les liens hiérarchiques entre concepts et de repérer la *distance sémantique* entre termes, par évaluation de leur position dans le corpus. La figure 20 montre l'application de la métrique de Rada à un thésaurus.

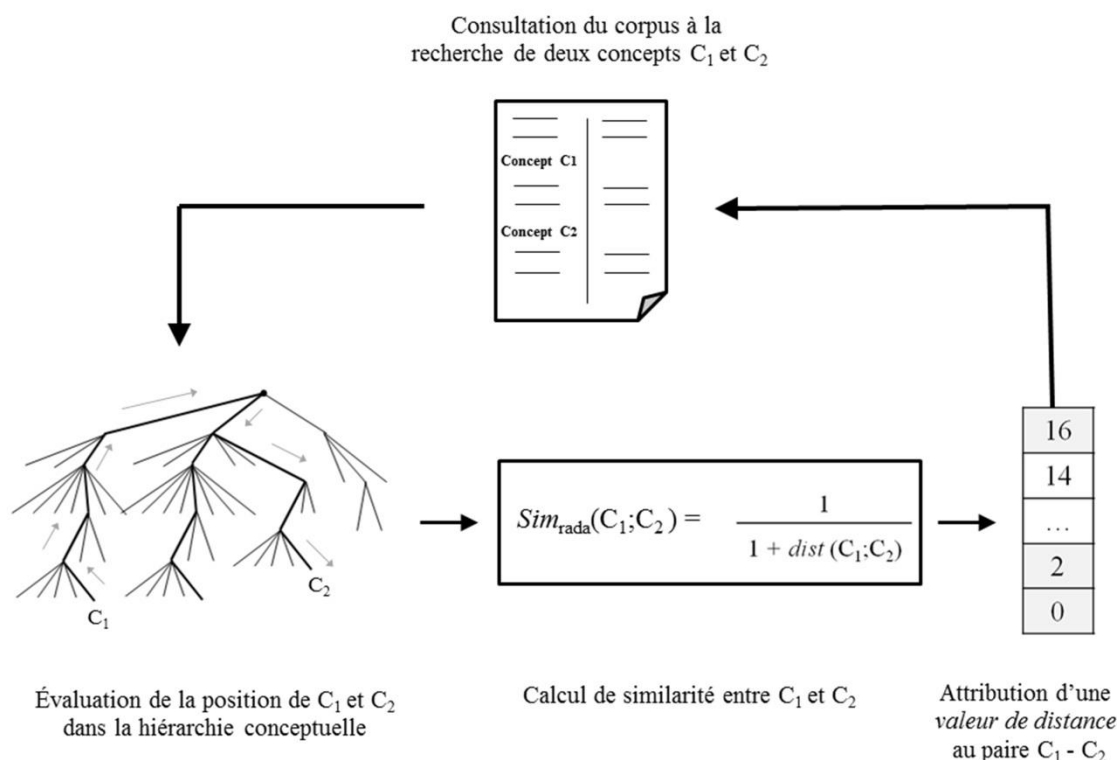
<sup>350</sup> *Axiome* : vérité évidente (terme mathématique).

<sup>351</sup> Formule cité et texte adapté d'après Maala et al (2007), p.10.

<sup>352</sup> Formule cité d'après Khatraoui et al. (2008), p.122 ; cf. aussi Maala et al . (2007), p.10.

<sup>353</sup> Cf. Khatraoui et al. (2008), p.123 ; cf. Ventresque et al. (2008), p.33.

<sup>354</sup> Formule cité d'après Ventresque et al. (2008), p.33 ; texte adapté de Ventresque et al. (2008) et Khataoui et al. (2008), p.122.



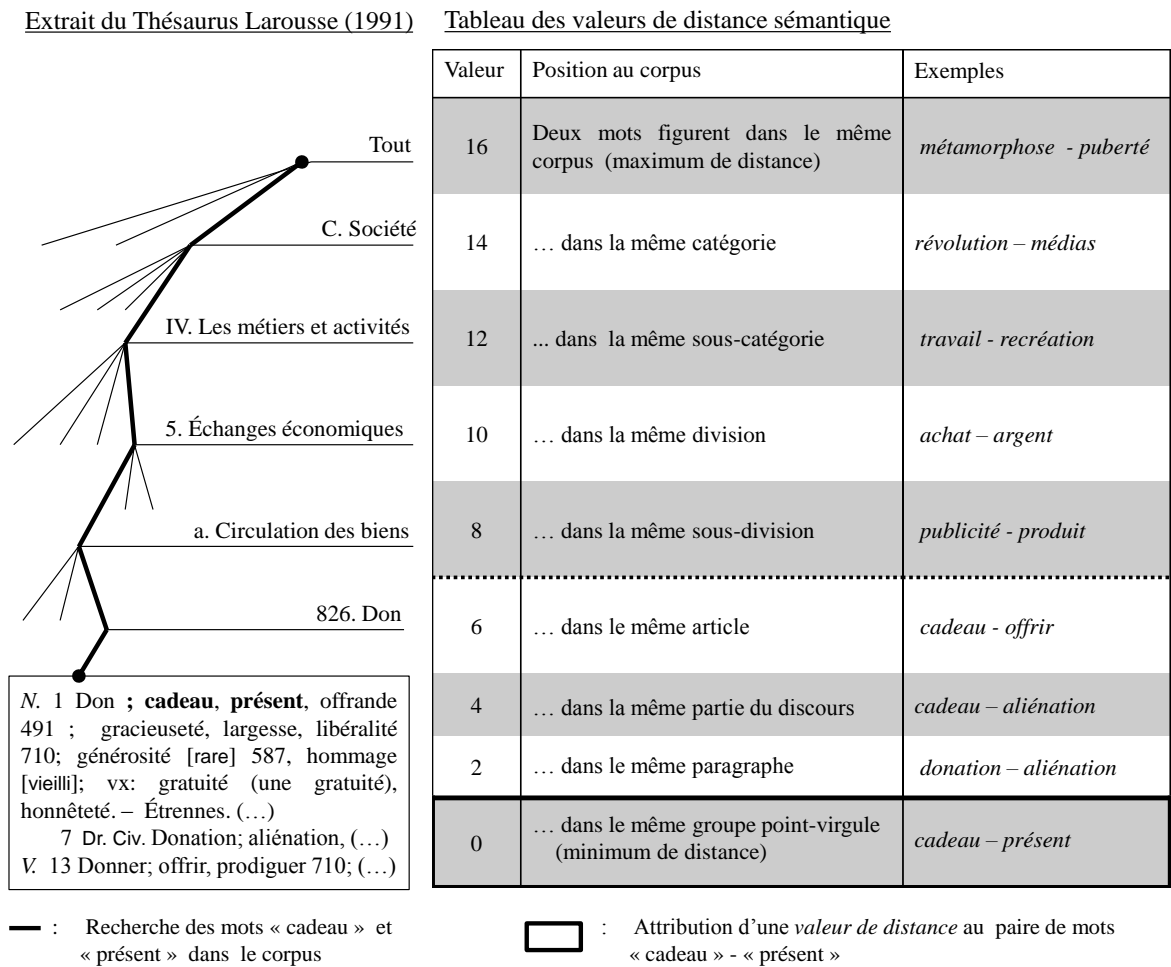
**Figure 20: Application de la mesure Rada à un thésaurus**<sup>355</sup>

Le schéma commence en haut : Lors de la consultation d'un thésaurus, l'utilisateur aperçoit deux termes qui lui apparaissent intéressant pour l'évaluation de similarité sémantique, par exemple '*achat*' et '*argent*'. Pour mesurer leur similarité, il recherche dans le corpus les concepts attachés aux termes, dans le cas d'exemple '*835. Dépense*' et '*839. Monnaie*'.<sup>356</sup> Ces concepts [appelés  $C_1$  et  $C_2$ ] dénotent la *notion* des entrées dictionnaires dans lesquels figurent les termes. Après avoir repérés les concepts, l'utilisateur prend l'aperçu de la hiérarchie des notions - imprimé sur les premières pages du thésaurus - pour chercher l'endroit où sont classés les concepts. L'objectif de sa mesure de similarité est d'évaluer la *distance sémantique* entre les concepts (auxquels sont attachés les termes). L'utilisateur s'appuie donc sur l'arborescence pour compter le nombre de liens (*arêtes*) entre les concepts (*nœuds*). La valeur évaluée [appelé  $dist(C_1;C_2)$ ] est introduite dans la métrique de Rada pour pouvoir attribuer une *valeur de distance* à la paire  $C_1 - C_2$ . La paire *achat - argent* se voit attribué, par exemple, une valeur de distance moyenne (10).

<sup>355</sup> Figure 20 de S.Reusch. Elaboré d'après les expériences faits lors de l'opérationnalisation de la mesure Rada pour le Thésaurus Larousse (1992).

<sup>356</sup> Cf. Thésaurus Larousse (1992), article 835. Dépense, p. 589 - pour le terme '*achat*'.  
Cf. Thésaurus Larousse (1992), article 839. Monnaie, p. 591 - pour le terme '*argent*'.

Une explication de la *mesure de distance sémantique* ne peut pas se limiter à une présentation d'un schéma d'application : une compréhension des *valeurs de distance* est aussi nécessaire. Elle seule permet de savoir, ce que nous comprenons par une '*distance moyenne*' dans le cas d'exemple de *achat - argent*. Les valeurs de distance sont déterminées par la structure hiérarchique du thésaurus dans lequel la mesure est opérée. Le Thésaurus Larousse repose sur 9 niveaux hiérarchiques auxquels des valeurs de distance sont attribués lors d'une mesure de distance sémantique.



**Figure 21 : Attribution des valeurs de distance aux concepts du Thésaurus Larousse**

La figure 21 montre un extrait de la *hiérarchie des notions* qui s'assortie d'un échantillon représentatif de valeurs de distance.<sup>357</sup> À droite, l'extrait de l'arborescence montre le

<sup>357</sup> Figure 21 de S.Reusch. Elaboré d'après les expériences faits lors de l'opérationnalisation de la mesure Rada pour le Thésaurus Larousse (1992). La mise en pratique de la mesure de Rada a été soutenu par une méthode proposé par Jarmasz pour la mesure de distance sémantique au sein du Thesaurus of English Words. Cf. Jarmasz (2003), pp. 38-42. Le tableau des valeurs a été crée avec des valeurs de distance sémantique proposé par Jarmasz (2003), p. 39 et a été remplis d'exemples concrets tirés des articles 190.

chemin que l'utilisateur descend - de la racine unique (indiqué par '*Tout*') à un article de dictionnaire (indiqué par 826. *Don*) - pour trouver deux termes parfaitement similaires (comme *cadeau* et *présent*). Sur le chemin vers le bas de la hiérarchie, l'utilisateur parcourt la totalité des niveaux hiérarchiques sur lesquels figurent d'autres paires de termes (comme *métamorphose* - *puberté* ou *publicité* - *produit*) qui ont une nature plus ou moins '*similaire*'. Ces exemples font partie d'un échantillon représentatif qui indique leur *valeur de distance*, évaluée sur la structure du Thésaurus Larousse. L'échantillon représentatif montre des valeurs de distance entre 0 et 16. Les distances reposent sur la *longueur du chemin* parcouru entre deux concepts, compté en nombres d'*arêtes*. La valeur 0 présente le minimum de distance. La valeur est attribuée à deux termes - tel que *cadeau* et *présent* - qui paraissent dans un même groupe point-virgule d'une entrée de dictionnaire. Le minimum de distance indique que deux termes partagent un sens *identique*, ils sont dits *synonymes*. Les valeurs entre 2 et 6 indiquent que deux termes - tel que *donation* - *aliénation* ou *cadeau* - *aliénation* - sont sémantiquement proches, car ils partagent un même concept ('*don*'). Les valeurs sont attribuées aux termes qui figurent - à différentes distances - dans une entrée de dictionnaire (826. *Don*). *Donation* et *aliénation* figurent dans le même paragraphe (2), *cadeau* et *aliénation* dans la même partie du discours - celle des noms (4) et les termes *cadeau* et *offrir* dans le même article (6). Les valeurs entre 8 et 14 montrent un accroissement de la distance sémantique - le chemin parcouru entre termes devient de plus en plus long. La distance sémantique s'accroît notamment lors de la transition vers le réseau sémantique - dans le schéma symbolisé par une ligne pointillée. À partir de cette ligne de démarcation, les termes appartiennent à deux entrées de dictionnaire différentes qui figurent - à différentes distances - dans le corpus du Thésaurus Larousse. Les termes *publicité* et *produit* figurent dans la même sous-division *a. Circulation des biens* (8), les termes *achat* et *argent* dans la même division *5. Échanges économiques* (10), les termes *travail* et *récréation* dans la même sous-catégorie *IV. Métiers et activités* (12) et les termes *révolution* et *médias* dans la même catégorie *C. Société* (14). La '*distance intuitive*' perçue lors de la considération des paires de mots (comme *révolution* - *médias*) dans un contexte *non-formalisé*, reçoit au sein du classement sémantique d'un thésaurus une valeur nominale : L'évaluation de la similarité sémantique devient ainsi une *mesure concrète*, la notion de similarité un *fait palpable* pour l'utilisateur. La valeur 16 représente

---

Évolution (p. 104), 315. Jeunesse (p. 219), 669. Politique (p.458), 726. Communication (p.498), 792. Emploi (p.550), 826. Don (p.582), 827. Commerce (p.583), 828. Marchandise (p.584), 835. Dépense (p.589), 839. Monnaie (p.591), 868. Passe-temps (p.620) du Thésaurus Larousse (1992).

le maximum de distance pour le Thésaurus Larousse. La valeur est attribuée à deux termes - tel que *métamorphose* et *puberté* - qui figurent sans liaison sémantique apparente dans le corpus du dictionnaire.<sup>358</sup> Le maximum de distance indique que deux concepts ont un sens *différent* et que leurs termes n'ont aucun lien sémantique apparent, qui repose sur la *structure classificatoire* du thésaurus. La notion de distance sémantique fortement marquée ou maximale, est une notion à utiliser avec précaution. Deux termes qui figurent à grande distance dans un thésaurus peuvent être rapprochés par une *relation d'association* (cf. chapitre 1.3.3). De voir la *puberté* comme la *métamorphose* de l'homme est un exemple que l'on crée lors d'un raisonnement par analogie.<sup>359</sup> L'appartenance des deux termes à différentes *classes conceptuelles* (celle de l' 'évolution' et celle de la 'jeunesse') fait que la mesure de similarité est remplacée, dans ce cas, par des considérations sur l'analogie.

Nous ne voulons pas terminer le chapitre sans une évaluation critique de la méthode. Pour le Thésaurus Larousse, le *calcul de distance sémantique* présente trois inconvénients. Le premier inconvénient est que la hiérarchie d'un thésaurus est basée sur des distances *uniformes*. Tous les liens sémantiques possèdent le même poids, ce qui pose problème pour la définition et le contrôle des distances des liens.<sup>360</sup> Le second inconvénient est que le calcul de similarité dépend de l'organisation des concepts dans la hiérarchie, adopté par un rédacteur.<sup>361</sup> L'organisation des articles peut avoir une influence sur la valeur de similarité.<sup>362</sup> L'exactitude du calcul dépend de l'intuition linguistique de constructeurs d'un thésaurus. Le troisième inconvénient est que le calcul présente une approche structurelle, qui ne tient pas compte de la description des concepts. Cela nous amène à tirer le bilan suivant : La méthode de l'*edge counting* est une mesure d'implémentation simple à l'étude d'un thésaurus.<sup>363</sup> D'un point de vue *structurel* elle a une bonne performance comparative par rapport à d'autres méthodes, mais pas d'un point de vue *informationnel*.<sup>364</sup> Ce fait nous amène à proposer une seconde méthode : La mesure basée sur la *proximité sémantique*.

<sup>358</sup> Cf. Jarmasz (2003), p. 38 ; cf. Jarmasz / Szpakowicz (2003), p. 215 ; cf. Kennedy / Szpakowicz, pp. 1/2.

<sup>359</sup> Informations tirés du Thésaurus Larousse (1992), Articles : 190. Évolution (p. 104) et 315. Jeunesse (p. 219).

<sup>360</sup> Cf. Slimani et al.(2007), p. 1.

<sup>361</sup> Cf. Khataoui et al. (2008), p. 123.

<sup>362</sup> Cf. Slimani et al.(2007), p. 1.

<sup>363</sup> Cf. Maala et al. (2007), p. 5.

<sup>364</sup> Cf. Maala et al. (2007), p. 6 ; cf. Slimani et al. (2007), p. 1.

### 3.5.2.2 Approche basée sur la proximité sémantique

La mesure basée sur la proximité sémantique prend en considération la *quantité d'informations* que partagent deux concepts (représentant les *nœuds d'un graphe*).<sup>365</sup> La méthode utilise un corpus d'apprentissage et mesure la probabilité de trouver un concept (appelé *concept subsumant*,  $C_s$ ) et ses descendants (appelés *concepts spécifiques*,  $C_1$  et  $C_2$ ) dans l'ouvrage.<sup>366</sup> Le concept générique  $C_s$  fournit la base pour la mesure de proximité sémantique entre deux concepts spécifiques  $C_1$  et  $C_2$ . Il 'subsume' leur quantité d'informations partagée, que l'on veut déterminer en vue d'un degré de similarité.<sup>367</sup> Ainsi, la pertinence du concept générique  $C_s$  dans le corpus sert à évaluer le *contenu informationnel* ( $C_i$ ) de deux concepts en comparaison.<sup>368</sup>

#### a) La mesure de Resnik

Philip Resnik [1995] propose une métrique qui intègre les données suivantes : Soit  $C_s$  un concept générique et  $C_i$  son contenu informationnel associé, alors  $p(C_s)$  sera la probabilité de retrouver une instance du concept  $C_s$  dans le corpus considéré.<sup>369</sup> On note<sup>370</sup> :

$$C_i(C_s) = -\log [p(C_s)]$$

Le thésaurus représente un corpus d'apprentissage spécifique, qui opère un classement des concepts au sein d'une hiérarchie des notions. La mesure de la proximité sémantique doit prendre en considération cette spécificité qui influence la nature du *concept subsumant*. Dans un thésaurus, le concept subsumant  $C_s$  est considéré comme le *concept le plus spécifique partagée* par deux concepts ( $C_1$  et  $C_2$ ) figurant dans la hiérarchie des notions.<sup>371</sup> Dans cette perspective, le concept subsumant est un concept partagé par  $C_1$  et  $C_2$ , qui se trouve par rapport à la racine d'une hiérarchie des notions au point le plus éloigné, donc le plus spécifique.<sup>372</sup> La métrique de Resnik tient compte de la notion de *concept le plus*

<sup>365</sup> Cf. Resnik (1999), p. 97 ; cf. Khataoui et al. (2007), p. 123.

<sup>366</sup> Cf. Khataoui et al. (2008), p.123.

<sup>367</sup> Cf. Resnik (1995), pp. 448/449 ; cf. Resnik (1999), p. 97 ; cf. Harrathi / Calabretto (2010), p.5 ; cf. Khataoui et al. (2008), p.123.

<sup>368</sup> Cf. Resnik (1995), pp. 448/449 ; cf. Resnik (1999), pp. 97( 98 ; cf. Khataoui et al. (2007), p.123 .

<sup>369</sup> Cf. Resnik (1995), p. 449 ; cf. Resnik (1999), p. 97 ; cf. Ventresque (2006), p. 373 ; cf. Ventresque et al. (2008), p.33 ; cf. Zargayouna / Salotti (2004), p. 4.

<sup>370</sup> Formule adopté et réarrangé d'après Resnik (1995), p. 449.

<sup>371</sup> Cf. Resnik (1995), p. 448 ; cf. Harrathi / Calabretto (2010), p.6 ; cf. Maala et al. (2007), p. 7.

<sup>372</sup> Cf. Maala et al. (2007), p. 7 ; cf. Ventresque (2006), p. 373 ; cf. Ventresque et al. (2008), p.33.

*spécifique partagée par deux concepts* et remplace la notation simple  $C_s$  par  $C_s(C_1, C_2)$  pour démontrer le lien existant.<sup>373</sup> On note<sup>374</sup> :

$$Sim_{resnik}(C_1; C_2) = -\log [p(C_s(C_1, C_2))]$$

Dans un corpus donné, la probabilité de retrouver le concept  $C_s$  se calcule par la fréquence totale du concept  $f$  qui se divise par la totalité des items figurant dans le corpus  $n$ .<sup>375</sup>

On note<sup>376</sup> :

$$p(C_s) = \frac{f(C_s)}{n}$$

La mesure de la similarité conceptuelle dans le Thésaurus Larousse est basée sur un chiffre approximatif  $n$  de 150.000 items qui figurent au total dans le corpus.<sup>377</sup> Le calcul se fait par un logarithme naturel (qualifié dans les notations suivantes par  $\log_n$ ).

La mesure de proximité sémantique nécessite un *corpus formalisé* pour opérer les calculs.<sup>378</sup> Le thésaurus, avec une organisation conceptuelle élaborée sur un consensus de plusieurs rédacteurs, fournit une base adéquate.<sup>379</sup> La hiérarchie de notions permet de repérer le *concept subsumant*  $C_s$  de deux concepts et l'index permet d'évaluer la *proximité sémantique* des termes attachés par la détermination de la *fréquence totale* du concept subsumant. La figure 22 montre un schéma d'application de la métrique de Resnik au Thésaurus Larousse qui est accompagné d'un exemple concret, la mesure de proximité sémantique pour le paire *donation* - *aliénation*.

<sup>373</sup> Cf. Resnik (1995), p. 449 ; cf. Resnik (1999), p. 97.

<sup>374</sup> Formule adaptée et réarrangée de Zargayouna / Salotti (2004), p.4.

<sup>375</sup> Cf. Resnik (1995), pp. 449/450 ; cf. Resnik (1999), pp. 99/100 ; cf. Zargayouna / Salotti (2004), p. 4 ; cf. Maala et al. (2007), p. 7 ; cf. Mc Hale (1998), p. 117.

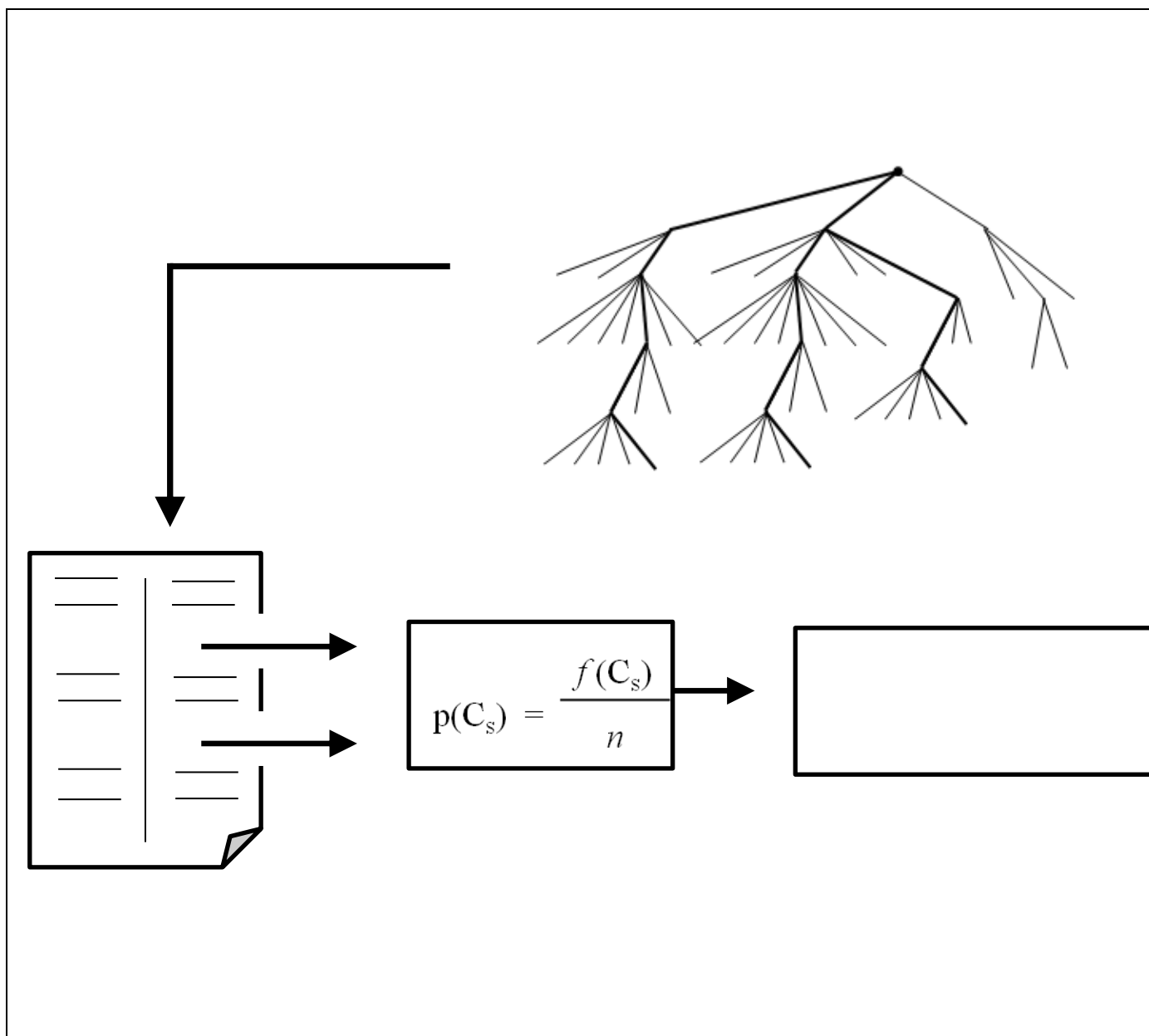
<sup>376</sup> Formule de S. Reusch adaptée et réarrangée de Resnik (1995), p. 449.

<sup>377</sup> Ce chiffre nous est donné par Thérèse Leridon, la responsable de presse de l'éditeur Larousse lors de la correspondance téléphonique du 27/05/2011.

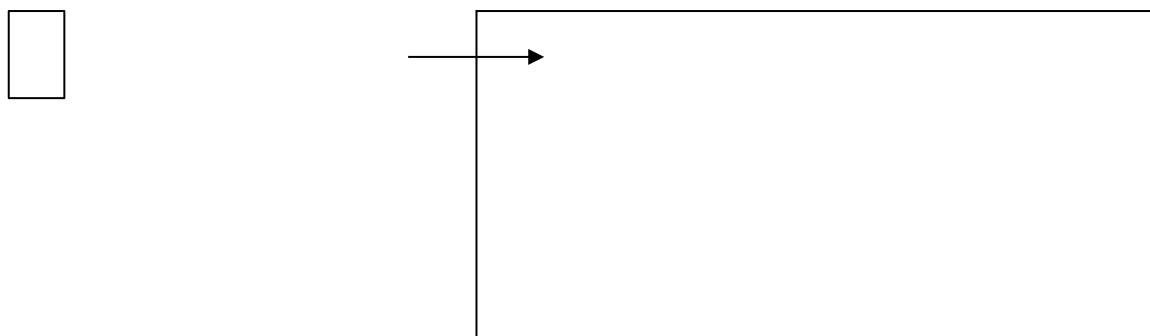
<sup>378</sup> Cf. Resnik (1995), p. 448.

<sup>379</sup> Cf. Ventresque (2006), p. 372.





Exemple : Calcul de similarité entre *donation* et *aliénation* (termes spécifiques de *don*)



## Figure 22 : Application de la mesure Resnik à un thésaurus<sup>380</sup>

Le schéma commence en haut : lors de la consultation du Thésaurus Larousse, l'utilisateur aperçoit deux termes - *donation* et *aliénation* - pour lesquels il veut opérer une mesure de similarité. Pour déterminer leur proximité sémantique, il recherche dans la hiérarchie des notions le *concept subsumant* de deux termes. *Donation* et *aliénation* figurent dans un même article de dictionnaire. Leur concept subsumant est le terme générique '*don*' qui dénote l'article 826. *Don* (cf. aussi figure 21).<sup>381</sup> Après l'avoir repéré, l'utilisateur cherche dans l'index des informations sur la probabilité de retrouver le concept '*don*' dans le corpus. L'entrée indiciaire montre 12 instances de '*don*' dans le Thésaurus Larousse.<sup>382</sup> Les instances sont attachées à plusieurs items (*don*, *cadeau*, *faveur*, *talent* etc.) qui dénotent plusieurs notions conceptuelles. Les différentes réalisations du concept - sous forme d'*item* - ne sont pas considérées lors d'une mesure de probabilité, seul le chiffre total d'occurrences est important. Ce chiffre 12 désigne la *fréquence* totale du concept subsumant '*don*' [notée  $f(C_s)$ ] qui est divisé par le chiffre total d'items qui figurent dans le Thésaurus Larousse [notée  $n$ ], donc par le chiffre 150.000. La *valeur de probabilité* [notée  $p(C_s)$ ] évaluée pour le concept subsumant 0.0008 est maintenant introduite dans la métrique de Rada pour obtenir la valeur de similarité de termes *donation* et *aliénation*. On note :  $\text{Sim}_{\text{resnik}}(\text{donation} ; \text{aliénation}) = -\log_n [0.00008]$ . La métrique de Resnik détermine une valeur de 9,43 pour la similarité sémantique entre les termes.

Les valeurs de proximité sémantique sont déterminées par le *contenu informatif* des entrées indiciaires d'un thésaurus. Il permet d'évaluer la probabilité de retrouver un concept dans le corpus, en fournissant sa *fréquence* compté en *nombre d'items*. Le Thésaurus Larousse présente la plus grande entrée indiciaire pour le concept '*temps*' (qui dénote en premier lieu l'article 170. *Temps*).<sup>383</sup> Il consiste en 139 *items*, le maximum de *fréquence* d'un terme.<sup>384</sup> La plus petite entrée indiciaire présente le Thésaurus Larousse pour le concept '*textile*' (qui dénote l'article 810. *Textile*).<sup>385</sup> Elle consiste en un seul *item*, celui du descripteur d'article.<sup>386</sup> Cette valeur constitue le minimum de *fréquence* d'un terme. La

<sup>380</sup> Figure 22 de S.Reusch. Elaboré d'après les expériences faits lors de l'opérationnalisation de la mesure Resnik pour la détermination de similarité sémantique entre les termes *donation* et *aliénation* au Thésaurus Larousse (1992).

<sup>381</sup> Cf. Thésaurus Larousse (1992), Corpus : article 826. Don, pp. 582/583.

<sup>382</sup> Cf. Thésaurus Larousse (1992), Index, p. 785. Sur cette page figure l'item « don ».

<sup>383</sup> Cf. Thésaurus Larousse (1992), article 170. Temps, pp. 92/93.

<sup>384</sup> Cf. Thésaurus Larousse (1992), Index, p. 1106. Sur cette page figure l'item « temps ».

<sup>385</sup> Cf. Thésaurus Larousse (1992), article 810. Textile, pp. 566/567.

<sup>386</sup> Cf. Thésaurus Larousse (1992), Index, p. 1110. Sur cette page figure l'item« textile ».

valeur du maximum et du minimum de *fréquence* sont introduits dans la métrique de Resnik pour obtenir de *valeurs limite* pour le calcul opéré dans le Trésor Larousse :

Maximum de fréquence d'un terme

$$p_{max}(Cs) = \frac{f(\text{temps})}{n}$$

$$p_{max}(\text{temps}) = \frac{139}{150.000} = 9,26666$$

$$= -\log_n [9,26666]$$

$$= 6,98391$$

Minimum de fréquence d'un terme

$$p_{min}(Cs) = \frac{f(\text{textile})}{n}$$

$$p_{min}(\text{textile}) = \frac{1}{150.000} = 6,66666$$

$$= -\log_n [6,66666]$$

$$= 11,91839$$

Dans le Thésaurus Larousse le maximum de fréquence d'un terme [noté 6,98] indique un degré de grande spécification en notions (cf. figure 22). Un terme, tel que '*temps*', est présenté dans toute sa variété conceptuelle (en tant que *instant*, *fil du temps*, *temps grammatical*, *temps pluvieux*, etc.). Par rapport à la grande quantité d'informations présente au corpus, deux termes tel que '*calendrier*' et '*chronologie*' montrent un minimum en informations partagées.<sup>387</sup> Ils partagent uniquement la notion de '*fil du temps*'. Dans le Thésaurus Larousse le minimum de fréquence d'un terme [noté 11,92] indique un degré de spécification indiscernable (cf. figure 22). Un terme, tel que '*textile*', est représenté par un seul *item*, de manière autoréférentielle. Par rapport à la quantité d'informations très réduite au corpus, deux termes tel que '*tissu*' et '*fil*' montrent un maximum en informations partagées.<sup>388</sup> Ils partagent l'unique notion existante, celle de '*textile*'. La définition du maximum et minimum en informations partagées, nous amène à déterminer la valeur de similarité pour les termes *donation* et *aliénation* [noté 9,43] comme représentant une grande proximité sémantique.

#### b) La mesure de Wu/Palmer

La mesure proposée par Wu et Palmer [1994] est basée sur l'exploitation d'une hiérarchie de termes dotée d'une racine unique, telle qu'elle existe dans un thésaurus.<sup>389</sup> La similarité entre deux concepts ( $C_1$  et  $C_2$ ) est calculée en comptant leur *distance* en nombre d'arêtes par rapport à leur concept subsumant  $C_s$  [noté  $dist(C_1, C_s)$  et  $dist(C_2, C_s)$ ] et en fonction de

<sup>387</sup> Cf. Thésaurus Larousse (1992), article 170.1 Temps, p. 92. Dans cet article figurent les deux items.

<sup>388</sup> Cf. Thésaurus Larousse (1992), article 810.1 Textile, p. 566. Dans cet article figurent les deux termes.

<sup>389</sup> Cf. Wu/Palmer (1994), pp. 133-138 ; cf. Ventresque et al. (2008), p.33.

la *profondeur* du concept subsumant dans la hiérarchie [noté  $prof(C_s)$ ].<sup>390</sup> Pendant le calcul, le concept subsumant fonctionne ainsi comme facteur de maximisation de la valeur sémantique partagée des deux concepts.<sup>391</sup> La méthode de Wu et Palmer se résume par la formule suivante :<sup>392</sup>

$$Sim_{wp}(C_1, C_2) = \frac{2 * prof(C_s)}{[dist(C_1, C_s) + dist(C_2, C_s)] + 2 * prof(C_s)}$$

Dans les schémas suivants nous utilisons une formule raccourcie<sup>393</sup>:

$$Sim_{wp}(C_1, C_2) = \frac{2 * N}{(N_1 + N_2) + 2 * N}$$

L'explication de la méthode Wu et Palmer montre que dans un thésaurus, la similarité de deux termes peut être définie par rapport à la *distance* et la *profondeur* de leurs concepts dans la hiérarchie. La *distance* est le chemin qui sépare les concepts de leur concept subsumant, en bas de la hiérarchie. La *profondeur* est le chemin qui sépare le concept subsumant de la racine unique, en haut de la hiérarchie.<sup>394</sup> La *distance* détermine la *position*, la *profondeur* la *spécification* des concepts (cf. figure 23-A). Pour mieux comprendre les notions de '*distance*' et '*profondeur*' de concepts, le schéma suivant propose une évaluation d'un exemple concret : La mesure de similarité entre les termes '*publicité*' et '*produit*' (cf. figure 23-B). Le schéma montre les concepts '*commerce*' et '*marchandise*' auxquels les termes sont attachés au sein du Thésaurus Larousse (cf. figure 21, valeur 8). Dans la hiérarchie des notions, ces concepts ont pour concept subsumant '*circulation des biens*'.

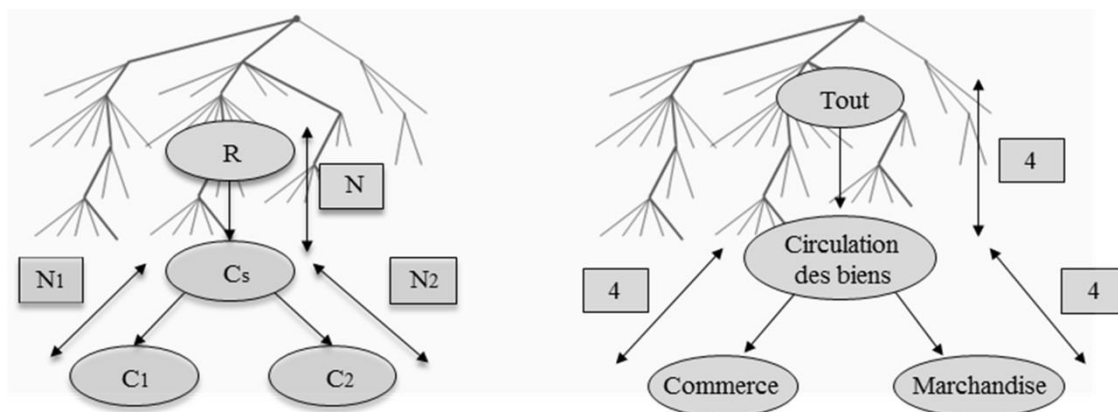
<sup>390</sup> Cf. Slimani et al. (2007), p. 3 ; cf. Zargayouna / Salotti (2004), p. 5 ; cf. Harrathi / Calabretto (2010), p. 5.

<sup>391</sup> Cf. Slimani et al. (2007), pp. 3/4.

<sup>392</sup> Formule citée d'après Harrathi / Calabretto (2010), p. 5 ; cf. aussi Wu / Palmer (1994), p. 136.

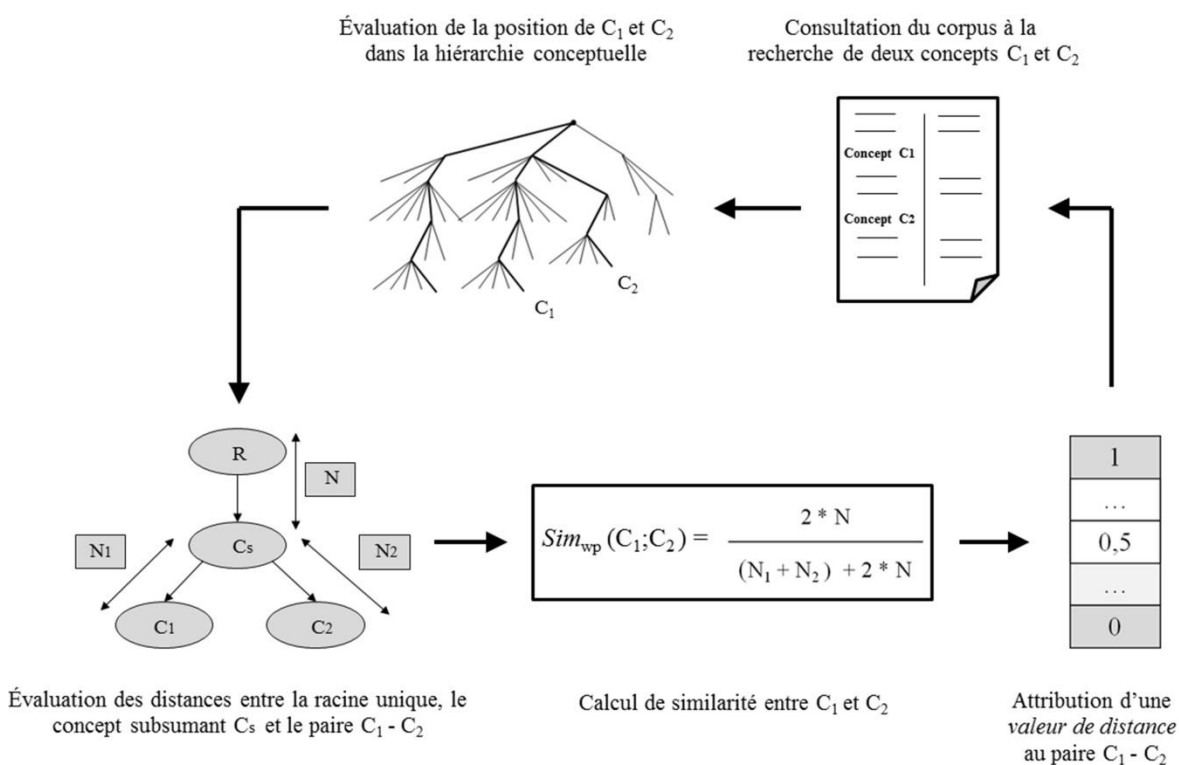
<sup>393</sup> Formule adoptée de Zargayouna / Salotti (2004), p. 5 ; cf. aussi Wu / Palmer (1994), p. 136.

<sup>394</sup> Cf. Slimani et al. (2007), p. 3 ; cf. Zargayouna / Salotti (2004), p. 5 ; cf. Ventresque (2006), p. 373 ; cf. Harrathi / Calabretto (2010), p.5.



**Figure 23: Schéma d'évaluation de distances proposées par Wu et Palmer**<sup>395</sup>

Lors de la consultation du Thésaurus Larousse les termes '*publicité*' et '*produit*' intéressent l'utilisateur en vue de similarité sémantique. Pour évaluer leur *proximité sémantique*, il recherche dans le corpus les concepts attachés aux termes (cf. figure 24). Dans notre cas d'exemple, '*commerce*' et '*marchandise*'.



**Figure 24: Schéma d'application de la mesure Wu/Palmer au Thésaurus Larousse**<sup>396</sup>

<sup>395</sup> Figure 23 de S.Reusch. Élaboré d'après un schéma proposé par Slimani et al. (2007), p.3 et avec des informations trouvés dans les articles 827. Commerce (p.583, pour le terme '*publicité*') et 828. Marchandise (p.584, pour le terme '*produit*') du Thésaurus Larousse (1992).

<sup>396</sup> Figure 24 de S.Reusch. Elaboré d'après les expériences faites lors de l'opérationnalisation de la mesure Wu/Palmer pour le Thésaurus Larousse (1992).

Après les avoir trouvés, il cherche dans la hiérarchie leur concept le plus spécifique partagé, leur « *concept subsumant* ». Ce concept '*subsume*' les propriétés de deux concepts.<sup>397</sup> Il représente donc une sorte de '*concept générique*' qui se situe à un niveau hiérarchique plus élevé que les concepts en considération (cf. figure 24). '*Circulation des biens*' est le concept subsumant de '*commerce*' et '*marchandise*' dans la hiérarchie du Thésaurus Larousse (cf. figure 23). Après l'avoir trouvé, l'utilisateur mesure la *profondeur* du concept subsumant, 'c'est-à-dire sa distance par rapport à la racine unique. Cette distance [notée  $N$  ou  $prof(C_s)$ ] a une valeur de 4 arêtes pour '*circulation des biens*'. Puis, il mesure la *distance* entre les concepts '*commerce*' et '*marchandise*' par rapport à leur concept subsumant. Ces distances [notées  $N_1$ ,  $N_2$  ou  $dist(C_1, C_s)$ ,  $dist(C_2, C_s)$ ] ont une valeur de 4 arêtes pour chacun des concepts (cf. figure 23). Maintenant, il reprend la formule proposée par Wu et Palmer,

$$Sim_{wp}(C_1, C_2) = \frac{2 * prof(C_s)}{[dist(C_1, C_s) + dist(C_2, C_s)] + 2 * prof(C_s)}$$

pour y introduire les valeurs évalués (cf. figure 24). Ainsi il peut calculer la similarité sémantique entre les termes '*publicité*' et '*produit*', de manière suivante :

Sim(Wu/Palmer): cadeau/présent

```
>> prof_Cs = 4;
>> dist_C1 = 4;
>> dist_C2 = 4;
>> sim = (2*prof_Cs)/(dist_C1,Cs+dist_C2,Cs)+(2*prof_Cs);
>> sim = (2*4)/(4+4)+(2*4)
sim = 0,5 .
```

Les termes '*publicité*' et '*produit*' ont été choisis comme premier exemple, parce qu'ils représentent la moyenne des valeurs de similarité sémantique pour la méthode Wu et Palmer [0 - 0,5 - 1] (cf. figure 24). On constate ainsi que leur valeur de similarité est suffisamment élevée pour détecter une proximité entre les deux termes [ $> 0$ ], tout en étant trop faible pour pouvoir constater une *équivalence sémantique* [ $< 1$ ]. Le tableau suivant explique les différentes valeurs de distances à l'aide d'exemples de termes qui figurent

<sup>397</sup> Cf. Harrathi / Calabretto (2010), p.6.

dans le Thésaurus Larousse, ces exemples ont été pré-utilisés déjà pour la méthode de Rada (cf. figure 21), ce qui permet une lecture comparative des différentes méthodes.<sup>398</sup>

**Tableau 21 : Valeurs de distance évaluées au Thésaurus Larousse**

Position au corpus	Exemples	Distances	Mesures	Valeurs
Deux mots figurent dans le même corpus (minimum de proximité)	<i>métamorphose - puberté</i>	N : 0 N <sub>1+2</sub> : 8	$\frac{2*0}{(8+8) + 2*0}$	0
...dans la même catégorie	<i>révolution - médias</i>	N : 1 N <sub>1+2</sub> : 7	$\frac{2*1}{(7+7) + 2*1}$	0,125
...dans la même sous-catégorie	<i>travail - récréation</i>	N : 2 N <sub>1+2</sub> : 6	$\frac{2*2}{(6+6) + 2*2}$	0,25
...dans la même division	<i>achat - argent</i>	N : 3 N <sub>1+2</sub> : 5	$\frac{2*3}{(5+5) + 2*3}$	0,375
...dans la même sous-division (proximité moyenne)	<i>publicité - produit</i>	N : 4 N <sub>1+2</sub> : 4	$\frac{2*4}{(4+4) + 2*4}$	0,5
...dans le même article	<i>cadeau - offrir</i>	N : 5 N <sub>1+2</sub> : 3	$\frac{2*5}{(3+3) + 2*5}$	0,625
...dans la même partie du discours	<i>cadeau - aliénation</i>	N : 5 N <sub>1+2</sub> : 2	$\frac{2*5}{(2+2) + 2*5}$	0,714
...dans le même paragraphe	<i>donation - aliénation</i>	N : 5 N <sub>1+2</sub> : 1	$\frac{2*5}{(1+1) + 2*5}$	0,833
...dans le même groupe point-virgule (maximum de proximité)	<i>cadeau - présent</i>	N : 5 N <sub>1+2</sub> : 0	$\frac{2*5}{(0+0) + 2*5}$	1

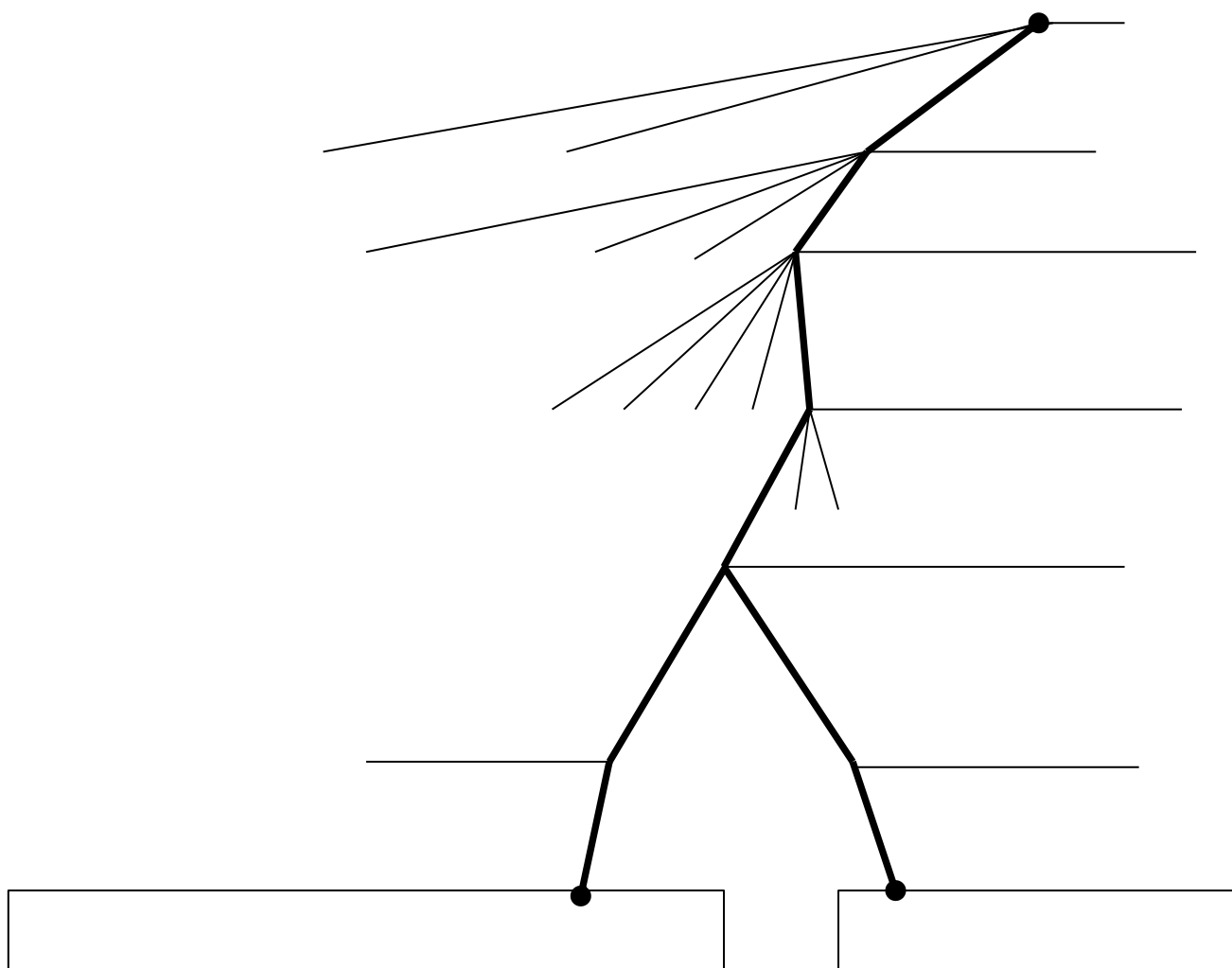
La hiérarchie des notions est un *système* qui permet de visualiser, par son classement sémantique, la *proximité sémantique* entre les termes d'un thésaurus. Dans le cas de *publicité - produit*, les articles, dans lesquels figurent les termes, sont organisés dans la même sous-division *a. Circulation des biens* (cf. figure 25). Cette sous-division consiste en six articles, 822. *Possession*, 823. *Cession*, 824. *Restitution*, 825. *Paiement*, 826. *Don*, 827.

<sup>398</sup> Tableau 21 de S. Reusch. Elaboré d'après les expériences faites lors de l'opérationnalisation de la mesure Wu/Palmer pour le Thésaurus Larousse (1992). La mise en pratique de la mesure de Rada a été soutenu par une méthode proposé par Jarmasz pour la mesure de distance sémantique au sein du Thesaurus of English Words. Cf. Jarmasz (2003), p. 38-42.

Le tableau des valeurs a été crée avec des valeurs de distance sémantique proposé par Jarmasz (2003), p. 23 et a été remplis d'exemples concrets tirés des articles 190. Évolution (p. 104), 315. Jeunesse (p. 219), 669. Politique (p.458), 726. Communication (p.498), 792. Emploi (p.550), 826. Don (p.582), 827. Commerce (p.583), 828. Marchandise (p.584), 835. Dépense (p.589), 839. Monnaie (p.591), 868. Passe-temps (p.620) du Thésaurus Larousse (1992).



*Commerce* et 828. *Marchandise*, dont les deux derniers sont représentés dans le schéma en bas.



**Figure 25: Calcul de similarité sémantique entre « publicité » et « produit »** <sup>399</sup>

<sup>399</sup> Figure 25 de S. Reusch. Élaboré d'après des informations trouvés dans le Thésaurus Larousse (1992), Sommaire, pp. xv-xxi et dans le Thésaurus Larousse (1992), Corpus : articles 827. Commerce (p.583) et 828. Marchandise (p.584).

La hiérarchie rapproche deux concepts, comme '*commerce*' et '*marchandise*', en vue d'une notion partagée : celle de la '*circulation des biens*'. Pourtant, deux termes comme '*publicité*' et '*produit*' ne se voient pas attribuer une valeur de similarité élevée uniquement à cause du rapprochement de leurs articles au sein d'un espace sémantique restreint : celui d'une sous-division. La mesure de Wu et Palmer prend en considération aussi le *degré de spécificité* de concepts. Il est indiqué par la *profondeur* à laquelle figurent les concepts, ainsi que leur notion attachée (dans notre cas la notion de '*circulation des biens*'), dans la hiérarchie. L'importance de la *profondeur* des concepts (d'un point de vue structurel) et l'importance de la *dépendance* des concepts d'un ancêtre spécifique commun (d'un point de vue informationnel) sont les deux éléments dans lesquels réside la particularité de la méthode Wu/Palmer. Une méthode qui se situe à la charnière entre l'*approche structurelle* [cf. Rada] et l'*approche informationnelle* [cf. Resnik] présentés dans ce chapitre. Ceci devient d'autant plus claire lors qu'on compare les différences [accentués en gras] entre la métrique de Wu/Palmer et celle de Rada :

$$\text{Sim}_{\text{rada}}(C_1;C_2) = \frac{1}{1 + \text{dist}(C_1 ;C_2)}$$

$$\text{Sim}_{\text{wp}}(C_1,C_2) = \frac{2 * \text{prof}(C_s)}{[\text{dist}(C_1, C_s) + \text{dist}(C_2, C_s)] + 2 * \text{prof}(C_s)}$$

À partir du premier exemple '*publicité*' - '*produit*', une série expérimentale peut être établie qui illustre le *rapprochement* (ou l'*éloignement*) des concepts à la racine unique du Thésaurus Larousse et qui montre par conséquent une *généralisation* (ou une *spécification*) des concepts testés en vue d'une similarité sémantique.

### 3.5.3 Considérations sur la mesure de similarité au sein d'un thésaurus

Dans le chapitre 3.5.2 nous avons présentés des méthodes pour l'évaluation de similarité sémantique dans un thésaurus. Les métriques utilisées permettent de formuler des hypothèses sur la taxonomie dans laquelle une mesure est opérée - dans notre cas sur le Thésaurus Larousse. Les métriques supposent, pour pouvoir opérer leur calcul, une

hiérarchie des termes qui repose sur un ordre conceptuel - la *hiérarchie des notions* du thésaurus présente une structure adéquate (cf. chapitre 1.3.4).<sup>400</sup>

La première approche de l'évaluation de similarité sémantique - examinée dans le chapitre 3.5.2.1 - est la *mesure basée sur la distance sémantique*. Cette mesure considère le thésaurus comme un '*arbre*' (math. un '*treillis*') qui figure dans un espace sémantique.<sup>401</sup>

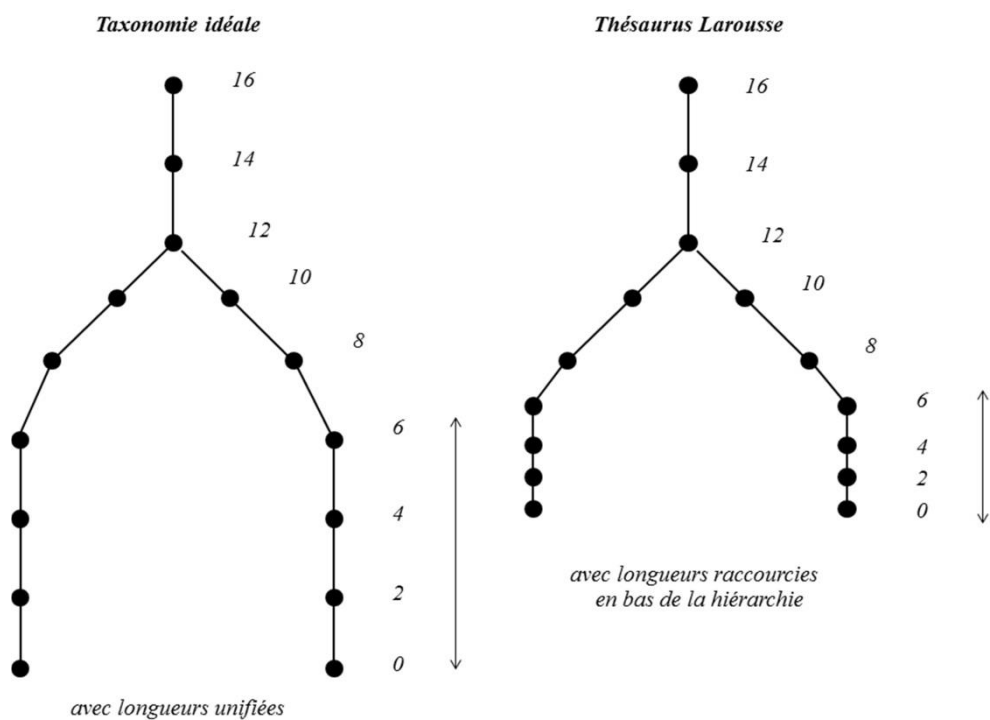
La *distance* entre concepts dans cet arbre - comptée en nombre d'*arêtes* entre *nœuds* - constitue le fondement pour la *méthode du comptage d'arêtes* (angl. *edge counting method*). La métrique proposée par Rada et al. (1989) suppose la méthode du comptage d'arêtes, qu'elle utilise pour déterminer un maximum et un minimum de distance relative à une taxonomie utilisée - le Thésaurus Larousse reçoit un spectre de valeur de 0 à 16. La *mesure de distance sémantique* opérée s'appuie - quant à elle - sur l'hypothèse de l'existence d'une taxonomie idéale qui repose sur des arêtes de longueur unifiée. Cette hypothèse peut constituer un problème pour la mesure de distance dans un thésaurus, car les arêtes y ont tendance à se raccourcir en profondeur de la hiérarchie - dans le Thésaurus Larousse ce phénomène concerne les arêtes au niveau des articles, en dessus d'une valeur de 8 (cf. figure 26).<sup>402</sup>

---

<sup>400</sup> Cf. Mc Hale (1998), p. 115.

<sup>401</sup> Cf. Mc Hale (1998), p. 115 ; cf. Aimé et al (2009), pp. 230/231.

<sup>402</sup> Cf. Mc Hale (1998), p. 115.



0 à 16: valeurs de distance pour la méthode de Rada et al. (1989)

**Figure 26: Hypothèse soulevée par Rada - une taxonomie idéale avec des longueurs unifiées**<sup>403</sup>

La mesure de distance ne prend pas en considération la '*densité des sous-hiérarchies*'. La '*densité*' s'explique comme suit : deux mots qui figurent dans une partie plus *dense* de la hiérarchie semblent plus étroitement liés que deux mots qui figurent dans une partie peu abondante.<sup>404</sup> Dans le Thésaurus Larousse, l'article 820. *Transports par air* contient beaucoup plus de mots que l'article 361. *Violet*. Cela ne signifie pourtant pas que '*avion*' est plus étroitement lié à '*hélicoptère*' que '*pourpre*' est lié à '*zinzolin*'.<sup>405</sup> La '*densité*' est un phénomène du classement sémantique qu'on devrait prendre en considération lors de l'évaluation de similarité sémantique dans un thésaurus - c'est un phénomène qui se produit lors d'un classement de mots, qui veut refléter au plus près possible la distribution des mots de la langue naturelle.<sup>406</sup> Une *mesure de distance sémantique* adéquate pour le Thésaurus Larousse devrait proposer une méthode qui utilise des *arêtes modifiées*, adaptées à la structure existante. Cette méthode devrait aussi associer un *poids* aux arêtes. Un poids pourrait prendre en considération le *type de lien* présent, la *profondeur du lien* dans le

<sup>403</sup> Figure 26 de S.Reusch. Élaboré d'après les expériences faits lors de l'opérationnalisation de la mesure Rada pour le Thésaurus Larousse (1992) et d'après les informations donnés par Mc Hale (1998), pp. 115.

<sup>404</sup> Cf. Agirre / Rigau (1996), pp.17 et 20.

<sup>405</sup> Cf. Thésaurus Larousse (1992), articles 361. Violet (p.258) et 820. Transports par air (p.578).

<sup>406</sup> Cf. Mc Hale (1998), p. 116.

thésaurus et de la *densité du concept* par rapport aux termes voisins immédiats.<sup>407</sup> Lors de la considération du *type de lien* on assignera un poids différent aux liens sémantiques (tels que la synonymie et l'antonymie), aux liens hiérarchiques (tels que l'hyponymie/l'hyperonymie) et aux liens associatifs (tel que l'analogie) entre termes. Lors de la considération de la *profondeur du lien* on assignera un poids différent aux liens entre '*superconcepts*' (tel que '*commerce*' et '*marchandise*' qui représentent deux descripteurs d'articles) et aux liens entre '*subconcepts*' (tel que '*négoce*' et '*produit*' qui représentent deux entrées dans un article).<sup>408</sup>

La deuxième approche de l'évaluation de similarité sémantique - examinée dans le chapitre 3.5.2.2 - est la *mesure basée sur la proximité sémantique*. Elle détermine la similarité sémantique entre deux concepts par la quantité de l'information qu'ils partagent - cette quantité est qualifiée comme le *contenu informationnel* de leur *concept subsumant*.<sup>409</sup>

La méthode de Resnik (1995) considère que le contenu informationnel du concept subsumant dépend de la '*probabilité*' de rencontrer une instance du concept dans le corpus.<sup>410</sup> La mesure s'appuie sur l'hypothèse qu'un thésaurus - comme le Thésaurus Larousse - montre une *fréquence* de concepts qui se rapproche aux valeurs évaluées dans d'autres textes généralisées.<sup>411</sup> Or, un thésaurus ne constitue pas un vocabulaire normalisé. La *probabilité* d'occurrences d'un concept dépend du nombre d'*items* qui figurent dans l'entrée indiciaire. Ce nombre dépend sensiblement de la *polysémie* des mots et d'autres conditions de *nature pragmatique*, comme le choix d'un rédacteur d'insérer un *item* dans plusieurs articles de dictionnaire.

La méthode de Wu et Palmer (1994) - présentée dans les dernières pages du chapitre précédent considère que le contenu informationnel du concept subsumant dépend de la '*profondeur*' à laquelle l'on trouve dans la hiérarchie. La méthode se concentre sur la *structure hiérarchique* du thésaurus - elle détecte la proximité sémantique entre deux concepts par rapport à leur distance au concept subsumant et par rapport à leur position dans la hiérarchie des notions.<sup>412</sup> La méthode s'appuie sur l'hypothèse que la similarité

<sup>407</sup> Cf. Zargayouna / Salotti (2004), pp. 8/9.

<sup>408</sup> Cf. Zargayouna / Salotti (2004), p. 9. Informations tirés du Thésaurus Larousse (1992), articles 826. Commerce (p.583) et 827. Marchandise (p. 584).

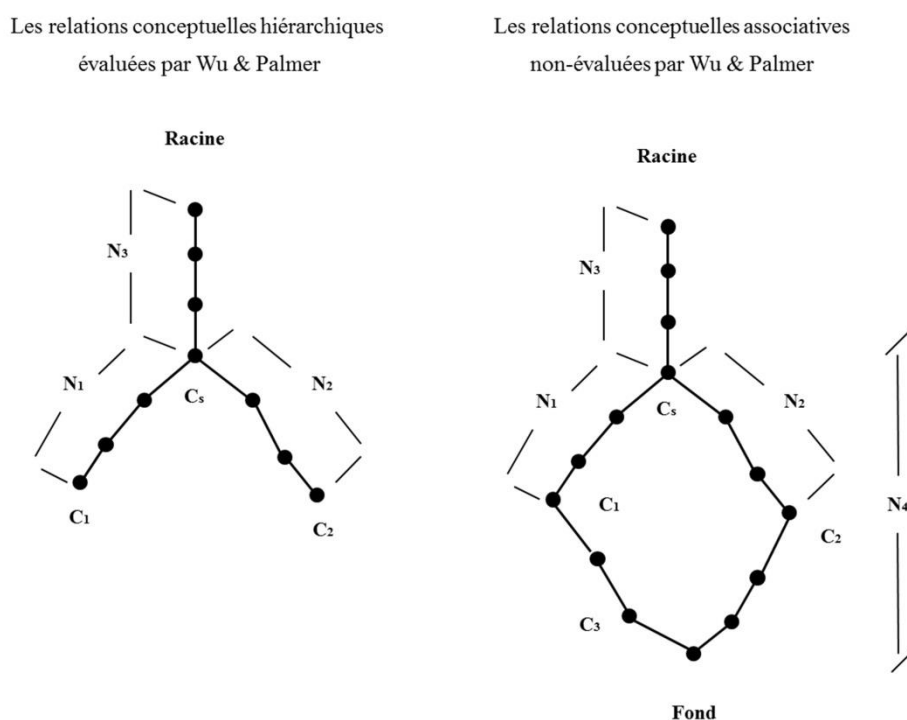
<sup>409</sup> Cf. Slimani et al. (2007), p. 4.

<sup>410</sup> Cf. Resnik (1995), pp. 448/449 ; cf. Zargayouna / Salotti (2007), p. 4.

<sup>411</sup> Cf. Mc Hale (1998), p. 116.

<sup>412</sup> Cf. Wu / Palmer (1994), p. 136 ; cf. Slimani et al. (2007), p. 6.

sémantique entre les termes d'un thésaurus, dépend essentiellement de la conception hiérarchique de l'ouvrage.<sup>413</sup> La métrique de Wu et Palmer évalue uniquement la position de concepts au sein de la hiérarchie [en utilisant les distances  $N_1$ ,  $N_2$  et  $N_3$ ] - ils ignorent les liens sémantiques qui s'instaurent dans la *partie invisible* de l'arborescence [évaluable par une quatrième distance,  $N_4$ ], au niveau du réseau sémantique (cf. figure 27).<sup>414</sup>



**Figure 27: Hypothèse de Wu/Palmer - une taxonomie qui se réduit à sa hiérarchie**<sup>415</sup>

La métrique de Wu et Palmer se restreint à l'évaluation des liens de *spécialisation/généralisation* et de *subordination/surordination* de concepts, même si l'évaluation d'autres relations non hiérarchiques - comme celle d'*association de termes* - peuvent aussi être important, par exemple, pour la mesure relationnelle d'une similarité conceptuelle symbolique.<sup>416</sup>

### 3.5.4 La synonymie : Un cas spécifique de similarité

<sup>413</sup> Cf. Zargayouna / Salotti ( 2004), p.8 ; cf. Slimani et al. (2007), p. 6.

<sup>414</sup> Cf. Zargayouna / Salotti (2004), p.6.

<sup>415</sup> Figure 27. Figures adoptés littéralement de Wu/Palmer (1994), p. 136 et de Zargayouna / Salotti ( 2004), pp. 5/6.

<sup>416</sup> Cf. Slimani et al. (2007), p. 3 ; cf. Zargayouna / Salotti (2004), pp. 8/9

La *synonymie* désigne la relation entre deux ou plusieurs *mots* (plusieurs signifiants) qui ont le même *sens* (un seul signifié).<sup>417</sup> On établit la synonymie par une procédure de *substitution* : Dans un même contexte, on remplace un mot par un autre.<sup>418</sup> Lorsque le sens de la proposition reste inchangé, les mots sont qualifiés de ‘synonymes’. La synonymie exprime ainsi une *équivalence sémantique* entre deux mots qui appartiennent à la même classe grammaticale.<sup>419</sup> Selon cette définition, il existe une équivalence sémantique entre ‘froid’ et ‘glacial’ quand on remplace<sup>420</sup> :

Le vent est *froid*.  
Le vent est *glacial*.

Néanmoins, si l’on examine les deux termes avec précaution, on s’aperçoit que la notion de *synonymie* est difficile à cerner avec rigueur. On ne peut substituer les deux termes dans tous les contextes.<sup>421</sup> Ainsi, on peut, avec les mêmes mots, construire une phrase où la synonymie disparaît<sup>422</sup> :

La soupe est *froide*.  
La soupe est ~~*glaciale*~~.

Les deux mots – *froid* et *glacial* – ont une acception identique pour certains contextes d’emploi, alors que dans d’autres contextes ils ont un sens divergent. L’exemple montre : Une différence doit être faite entre les *synonymes absolus* et des *quasi-synonymes*, de valeur approximative. Les *synonymes* sont deux mots qui sont situables dans n’importe quel contexte. Les *quasi-synonymes* sont deux mots qui commutent uniquement dans un contexte déterminé.<sup>423</sup>

Dans un thésaurus, la *synonymie* est une relation sémantique centrale. Sur le plan structurel, elle joue un rôle important dans le classement sémantique. Elle organise les *mots d’entrée* du dictionnaire, en bas de la hiérarchie des notions (cf. chapitre 3.5.2). Sur le plan sémantique, elle joue un rôle important dans la recherche d’informations. Elle décrit la *relation d’équivalence sémantique entre termes*, dans une entrée en considération (cf.

<sup>417</sup> Cf. Schwab (2005), p. 93 ; cf. Old (2003), pp. 12-16 ; cf. Lehmann /Martin-Berthet (2000), p. 54.

<sup>418</sup> Cf. Dictionnaire de linguistique (1994), Article : Synonymie, p. 465 ; cf. Niklas-Salminen (1997), p. 110 ; cf. Lamizet (1975), pp. 17/18 ; cf. Schwab et al (2003), p.240 ; cf. Schwab (2005), p. 93.

<sup>419</sup> Cf. Niklas-Salminen (1997), p. 110. ; cf. Schwab et al (2003), p.240 ; cf. Lehmann /Martin-Berthet (2000), p. 54.

<sup>420</sup> Exemple de S. Reusch

<sup>421</sup> Cf. Dictionnaire de linguistique (1994), Article : Synonymie, p. 466 ; cf. Niklas-Salminen (1997), p. 110.

<sup>422</sup> Exemple de S. Reusch

<sup>423</sup> Cf. Niklas-Salminen (1997), p. 111 ; cf. Polguère (2003), pp. 122-124.



chapitre 1.3.3).<sup>424</sup> En tant qu'*élément de classification*, la synonymie circonscrit un champ d'intervention précis dans le thésaurus. En tant qu'*élément de sémantique*, la synonymie décrit un rapport entre termes, dont la nature reste floue. Pour pouvoir constater une *relation d'équivalence*, la synonymie devrait montrer les qualités de symétrie, de réflexivité et de transitivité – or ces propriétés ne sont pas toujours vérifiables.<sup>425</sup> Prenons l'exemple de quelques mots classés comme '*synonymes*' au sein du Thésaurus Larousse.<sup>426</sup>

1. Le premier exemple montre que la relation entre synonymes n'est pas toujours *transitive*.<sup>427</sup> L'entrée 32 *Innovation* indique les verbes *concevoir*, *imaginer* et *élaborer* comme synonymes. Les verbes '*concevoir*' et '*imaginer*' se substituent. C'est-à-dire, qu'ils expriment dans plusieurs contextes d'usage, de façon identique le concept '*penser l'innovation*'.<sup>428</sup> Leur relation est dite *transitive*. Les verbes '*concevoir*' et '*élaborer*' ne se substituent pas. Le premier verbe indique un procès mental ('penser l'innovateur'), le second un procès physique ('faire l'innovateur'). Leur relation est dite *intransitive*. Les *signifiants* d'un mot ne sont pas transférable dans un autre contexte pour exprimer le *signifié* un autre mot. L'exemple montre qu'en vue de la classification d'un thésaurus, « *la synonymie est au mieux une relation de tolérance* ». <sup>429</sup>

2. Le second exemple montre que la relation entre synonymes n'est pas toujours *symétrique*.<sup>430</sup> Dans un thésaurus, la synonymie peut se confondre avec la notion de hyperonymie/hyponymie – une relation *hiérarchique*.<sup>431</sup> Dans l'entrée 848 *Maison* le substantif '*maison*' a pour synonymes '*demeure*', '*logement*' et '*résidence*' alors qu'il en est hyperonyme.<sup>432</sup> Le substantif '*maison*' fonctionne comme régissant au paragraphe 848.1 où figurent les autres substantifs. Il représente le terme le plus neutre et le plus usité, par rapport aux autres.<sup>433</sup> De plus, il s'imbrique – en position du premier mot de l'entrée –

<sup>424</sup> Cf. Lafourcade/Prince (2001), p. 233 ; cf. Gouvernement du Québec (2011), Politique d'édition du TAG, pp. 8/9 ; cf. Carlisle (2011), Guidance Manual, p. 7.

<sup>425</sup> Cf. Lafourcade/Prince (2001), p. 233.

<sup>426</sup> „classés comme synonymes“ signifie que l'on classe deux verbes, l'un à côté de l'autre, au sein d'un même groupe point-virgule d'une entrée de dictionnaire (cf. chapitre 3.1.2.1, notamment figure 21 et la suite du présent chapitre)

<sup>427</sup> Cf. Lafourcade/Prince (2001), p. 234.

<sup>428</sup> Exemple tiré du Thésaurus Larousse (1992), article 32. Innovation, pp.16/17.

<sup>429</sup> Cité d'après Lafourcade/Prince (2001), p.234.

Lafourcade/Prince considèrent qu'une relation de tolérance est symétrique et réflexive mais *pas transitive*.

<sup>430</sup> Cf. Lafourcade/Prince (2001), p. 234.

<sup>431</sup> La relation de hyperonymie/hyponymie s'instaure dans un thésaurus entre le « concept verbal » - le premier verbe, le plus neutre en considération et les verbes suivants, plus spécifiés qui figurent dans la même entrée.

<sup>432</sup> Exemples tiré du Thésaurus Larousse (1992), article 848. Maison, p. 600.

<sup>433</sup> Cf. Lafourcade/Prince (2001), p. 234.

avec le *mot concept* ‘*maison*’, qui dénote la notion sémantique de tout mot et de toute locution, figurant dans l’entrée. La position hyperonyme de ‘*maison*’ fragilise la notion de *symétrie* et anéantit ainsi l’acception *d’une relation d’équivalence* entre les substantifs en considération. Ainsi, on considère l’hyperonyme ‘*maison*’ comme synonyme de ‘*demeure*’, ‘*logement*’ et ‘*résidence*’, parce qu’il partage avec eux toutes les propriétés.<sup>434</sup> Inversement, sur le plan sémantique, les hyponymes ne sont pas considérés comme synonymes de ‘*maison*’.<sup>435</sup> Ainsi, ‘*demeure*’ n’apparaît pas comme synonyme de ‘*maison*’, alors que l’inverse s’avère admissible.<sup>436</sup>

3. Le troisième exemple montre que la relation entre synonymes n’est pas toujours *réflexive*.<sup>437</sup> Deux hyponymes d’un même terme, malgré une forte parenté, ne sont pas obligatoirement synonymes.<sup>438</sup> Les rapports de *quasi-synonymie* sont beaucoup plus fréquents dans le classement d’un thésaurus. Dans l’entrée 699 *Vertu* les adjectifs ‘*innocent*’ et ‘*immaculé*’ sont définis comme les synonymes de ‘*chaste*’, qui est leur *terme générique commun*.<sup>439</sup> Malgré leur origine sémantique commune, qui revoit à la notion de *vertu de l’homme*, les termes ne montrent pas les qualités de synonymes à cause d’une connotation sémantique très forte qui change dans le contexte d’emploi :

Contexte 1 : synonymie

Marie était une fille *chaste*, jamais un homme n’avait le droit de toucher à elle.

Marie était une fille *innocente*, jamais un homme n’avait le droit de toucher à elle.

Contexte 2 : quasi-synonymie

Marie était une âme *innocente*, jamais elle avait volé quelque chose.

Marie était une âme ~~*innocente*~~ *immaculée*, jamais elle avait volé quelque chose.

Cerner la nature de synonymie dans le cadre du classement sémantique d’un thésaurus est un exercice difficile. La définition de la synonymie comme ‘relation qui s’instaure entre deux items du même groupe point-virgule d’une entrée de dictionnaire’, telle qu’elle est

<sup>434</sup> Un *terme générique* (ou *descripteur*) – tel que le substantif ‘*maison*’ – lègue tous ses propriétés sémantiques (dits propriétés génériques) aux termes spécifiques – tel que le substantif ‘*demeure*’.

<sup>435</sup> Un *terme spécifique* (ou *mot d’entrée*) – tel que le substantif ‘*demeure*’ – ne peut pas léguer ses propriétés sémantiques (dits propriétés spécifiques) à un terme générique – tel que le substantif ‘*maison*’. Il peut uniquement recevoir ses propriétés et se définir ainsi par rapport à celui-ci. Ce schéma héréditaire explique la dissolution de la symétrie : L’héritage des propriétés sémantique s’opère de façon unidirectionnelle.

<sup>436</sup> Cf. Lafourcade/Prince (2001), p. 234.

<sup>437</sup> Cf. Lafourcade/Prince (2001), p. 234.

<sup>438</sup> Cf. Lafourcade/Prince (2001), p. 234.

<sup>439</sup> Exemple tiré du Thésaurus Larousse (1992), article 699. Politique, p. 478.

indiquée par la *mesure de similarité*, est trop vague. Même si la définition résulte d'une analyse structurale, elle ne tient pas compte du caractère constructif des *chaînes sémantiques* au sein d'un thésaurus. La composition d'un article reflète avant tout un choix linguistique fait par un à dix rédacteurs (au maximum !). En outre, l'emploi du point-virgule comme seule notation d'équivalence sémantique fait remarquer la non-délimitation entre synonymie et quasi-synonymie dans un corpus comme le Thésaurus Larousse. Ceci montre la fragilité de la définition de *synonymie* sur le plan structurel. Le mélange constant entre *synonymie* et *quasi-synonymie* – au sein des termes qui figurent dans un même groupe point-virgule – crée, à la hauteur du corpus, une *relation d'équivalence entre termes* qui se caractérise par un changement perpétuel des propriétés – de symétrie, transitivité et réflexivité.<sup>440</sup> Ceci montre la fragilité de la définition de *synonymie* sur le plan sémantique. Sans délimitation précise entre ces deux notions, l'utilisateur d'un thésaurus est seul devant le choix du bon terme dans un ensemble de '*distracteurs*', qui figurent dans le même environnement sémantique. Ceci est pour un utilisateur d'origine étrangère très déconcertant.<sup>441</sup> L'utilisateur étranger n'aperçoit guère les différences plus ou moins sensibles qui existent entre les *signifiants* pour exprimer un même *signifié* dans un *contexte d'usage*.<sup>442</sup> La seule chose qui lui restera, sera de croire, à tout moment, d'être tombé sur un cas de *synonymie absolue* – même s'il conçoit sa rareté.<sup>443</sup>

Une délimitation claire entre *synonymie* et *quasi-synonymie* est un problème lexicographique constant qui se retrouve dans les ressources lexicales électroniques modernes, comme WordNet. Le clivage entre le *contexte formel* du dictionnaire et le *contexte d'usage* de l'utilisateur s'y manifeste d'avantage.

## 4. Le Thésaurus numérique - Enjeux, potentiels et perspectives

### 4.1 WordNet et FrameNet - Force et faiblesse des ressources numériques

#### 4.1.1 WordNet : Un système de classement à base de *synsets*

WordNet est une ressource lexicale développée par les linguistes du laboratoire des sciences cognitives de l'université de Princeton. Le système se présente sous la forme

<sup>440</sup> Cf. Lafourcade/Prince (2001), pp. 233/234.

<sup>441</sup> Nous considérons par un '*distracteur*', un terme qui est dit *synonyme* d'un terme générique, mais qui est pour l'utilisateur sans rapport à son contexte d'utilisation désiré.

<sup>442</sup> Cf. Batteux (2000), p. 105 ; cf. Lafourcade/Prince (2001), pp. 234/235.

<sup>443</sup> Cf. Niklas-Salminen (1997), p. 111 ; cf. Schwab (2005), p. 93 ; cf. Polguère (2003), p. 122.

d'une base de données électronique qu'on peut consulter en ligne ou télécharger sur un système local.<sup>444</sup> Il est inspiré des théories psycholinguistiques sur la mémoire lexicale humaine (cf. Collins/Quillian 1969). Son objectif est de répertorier, classer et mettre en relation de diverses façons le contenu sémantique de la langue anglaise.<sup>445</sup> WordNet regroupe les mots en ensembles de synonymes cognitifs (*synsets*) qui expriment chacun une notion ou un usage particulier d'un concept.<sup>446</sup> Dans WordNet 2.1 le nom '*crime*' est défini par deux *synsets* :

1. **crime**, law-breaking -- ((criminal law) an act punishable by law; usually considered as an evil act; "a long record of crimes")
2. **crime** -- (an evil act not necessarily punishable by law; "crimes of the heart")

Chaque *synset* dénote une acception différente du nom '*crime*', décrite par une courte définition. Tout comme un dictionnaire analogique classique, WordNet offre ainsi pour chaque mot, une liste de synonymes, correspondant à toutes ses acceptions répertoriées.<sup>447</sup> Les *synsets* sont placés dans une structure arborescente qui est hiérarchisée par la notion d'hyponymie.<sup>448</sup> Un exemple de hiérarchie de *synsets*, évoqué par l'application HYPERNYMS, est donné par le tableau suivant pour le *synset* '*crime, law-breaking*'<sup>449</sup>:

entity
abstract entity
abstraction
psychological feature
event
act, human action, human activity
activity
wrongdoing, wrongful conduct, misconduct, actus reus
transgression, evildoing
<i>crime, law-breaking</i>

WordNet est construit comme un graphe conceptuel, qui est constitué de *nœuds* (concepts) et *arêtes* (liens entre concepts). Les *synsets* sont reliés entre eux par des relations sémantiques, autres que celle de *synonymie*. Comme le Thésaurus Larousse, WordNet

<sup>444</sup> Cf. Chaumartin (2007), p. 60 ; cf. Fellbaum et al. (1998), pp. 111/112.

<sup>445</sup> Cf. Fellbaum et al. (1998), pp. 31/32.

<sup>446</sup> Cf. Fellbaum (1999), pp. 27/28 ; cf. Chaumartin (2007), p. 60 ; cf. Stuckenschmidt (2009), pp. 55-58 ; cf. Safar et al. (2007), p. 26.

<sup>447</sup> Cf. Fellbaum et al. (1998), pp. 24-26.

<sup>448</sup> Cf. Chaumartin (2007), p. 60 ; cf. Stuckenschmidt (2009), pp. 58-60.

<sup>449</sup> Cf. WordNet 2.1 pour l'entrée « crime », consultable sur : <http://wordnet.princeton.edu/>.

répertorie une grande variété de relations sémantiques, qui permet d'organiser le sens des mots en systèmes de catégories qu'on peut consulter de manière cohérente.<sup>450</sup> Ainsi, on peut interroger le système, par l'application HYPONYMS, aux *synsets* qui présentent les différents scénarios de '*crime*'<sup>451</sup>:

HYPONYMS : *fraud*

HYPONYMS : *warcrime*

HYPONYMS : *treason, high treason, lese majesty*

On peut également interroger le système, par l'application COORDINATE TERM, aux *synsets* présentant les notions conceptuelles proche à celle de '*crime*'<sup>452</sup>:

COORDINATE-TERM : *abomination*

COORDINATE-TERM : *vice*

COORDINATE-TERM : *evil, immorality, wickedness, iniquity*

... ou par l'application TOPIC TERM, aux *synsets* présentant des expressions verbales utilisés pour décrire un acte criminel<sup>453</sup> :

TOPIC TERM : *to kidnap*

TOPIC TERM : *to blackmail*

TOPIC TERM : *to shoplift*

L'analyse montre que les *synsets* sont reliés entre eux par des relations conceptuelles, sémantiques et lexicales diverses.<sup>454</sup> Cependant, la couverture de sémantique lexicale dans WordNet n'est pas exhaustive. Dans sa conception, WordNet ne tient pas compte de la *sémantique de discours*, qui indique un *contexte conceptuel* dans lequel est inscrit un environnement syntagmatique qui réalise des mots attachés à une topique de discours.<sup>455</sup> Par exemple, les mots '*raquet*', '*ball*' et '*net*' ne sont pas étroitement liés pour former un concept '*court game*'. Dans la monographie « WordNet: an Electronic Lexical Database » (1998) Christiane Fellbaum surnomme cette particularité '*the tennis problem*' et l'explique de manière suivante :

---

<sup>450</sup> Cf. Stuckenschmidt (2009), pp. 58-65 ; ; cf. Fellbaum et al. (1998), pp. 9/10.

<sup>451</sup> Cf. WordNet 2.1 pour l'entrée « crime », consultable sur : <http://wordnet.princeton.edu/>.

<sup>452</sup> Cf. WordNet 2.1 pour l'entrée « crime ».

<sup>453</sup> Cf. WordNet 2.1 pour l'entrée « crime ».

<sup>454</sup> Cf. Stuckenschmidt (2009), pp. 58-65 ; cf. Veale (2006), pp. 105-110 ; cf. Fellbaum et al. (1998), pp. 9/10.

<sup>455</sup> Cf. Chaumartin (2007), p. 64 ; cf. Fellbaum et al. (1998), pp. 10/11.

“WordNet links words and concepts through a variety of semantic relations based on similarity and contrast. Because it focuses on the semantics of words and concepts rather than on semantics at the text or discourse level, WordNet contains no relations that indicate the word’s shared membership in a topic of discourse: For example, WordNet does not link *racquet*, *ball*, and *net* in a way that would show that these words, and the concepts behind them, are part of another concept that can be expressed by *court game*.”<sup>456</sup>

Une ressource électronique, qui ignore un rapprochement de termes en vue d’un *contexte* déterminé par une communauté d’usage, cache des renseignements potentiellement utiles pour la recherche d’informations.<sup>457</sup>

#### 4.1.2 FrameNet : Un système de classement à base de *frames*

FrameNet est une ressource lexicale qui s’articule autour de la notion de *rôle thématique*.<sup>458</sup> Suivant la *Sémantique des cadres* de Charles Fillmore, le sens est représenté à partir des *cadres conceptuels* (angl. *frames*) qui peuvent être considérés comme de représentations sémantiques d’objets, de situations ou d’événements.<sup>459</sup> L’exemple suivant montre, par alignement à WordNet, le *frame* pour ‘crime’<sup>460</sup>:

*frame : crime\_scenario*

A (putative) Crime is committed and comes to the attention of the Authorities. In response, there is a Criminal\_investigation and (often) Arrest and criminal court proceedings. The Investigation, Arrest, and other parts of the Criminal\_Process are pursued in order to find a Suspect (who then may enter the Criminal\_process to become the Defendant) and determine if this Suspect matches the Perpetrator of the Crime, and also to determine if the Charges match the Crime. If the Suspect is deemed to have committed the Crime, then they are generally given some punishment commensurate with the Charges.

Dans un *frame* les rôles sémantiques sont appelés *frame elements*. Ils sont attachés de manière unique à chaque *frame*.<sup>461</sup> Le *frame* ‘crime\_scenario’ englobe cinq rôles sémantiques<sup>462</sup> :

<sup>456</sup> Cité d’après Fellbaum et al. (1998), p.10.

<sup>457</sup> Cf. Aimé et al. (2009), p. 238.

<sup>458</sup> Cf. Baker (2009), p. 32.

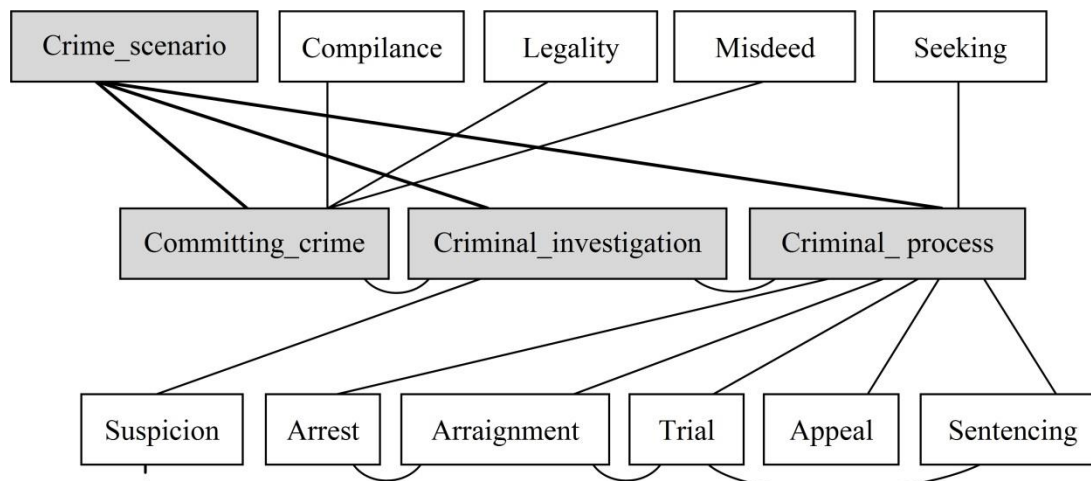
<sup>459</sup> Cf. Pado / Pitel (2007), p. 2 ; cf. Steiner (2003), p. 119.

<sup>460</sup> Cf. FrameNet consultable sur : <https://framenet.icsi.berkeley.edu> ; cf. Chaumartin (2007), p. 68.

<sup>461</sup> Cf. Chaumartin (2007), p. 69 ; cf. Pado / Pitel (2007), p. 2 ; cf. Steiner (2003), p. 120 ; cf. Fillmore et al. (2001), p. 9.

Authorities	The group which is responsible for the maintenance of law and order, and as such have been given the power to investigate <u>Crimes</u> , find <u>Suspects</u> and determine if a Suspect should be submitted to the Criminal_Process.
Charge	A description of a type of act that is not permissible according to the law of society.
Crime	An act, generally intentional, that matches the description that belongs to an official <u>Charge</u> .
Perpetrator	The individual that commits a <u>Crime</u> .
Suspect	The individual that is under suspicion of having committed a <u>Crime</u> .

La conception de FrameNet consiste à construire une ressource lexicale qui met en relation les *frames*, les lemmes qui les évoquent (appelés *lexical units*) et les informations détaillées sur les relations de surface des *frames éléments* qui prennent la forme des cadres syntaxiques et grammaticaux avec un ensemble des phrases d'exemples annotés sur divers dictionnaires anglais.<sup>463</sup> Le schéma suivant montre les cadres associés à 'crime\_scenario' <sup>464</sup> :



- Martin McGuinness, a leader of Sinn Fein, was arrested in Dublin last night on suspicion of being a member of an unlawful organisation. (CNI)

<sup>462</sup> Cf. FrameNet pour l'entrée « crime scenario », consultable sur : <https://framenet.icsi.berkeley.edu> .

<sup>463</sup> Cf. Pado / Pitel (2007), p. 2 ; cf. Steiner (2003), pp. 120-122 ; cf. Fillmore et al. (2001), p. 9.

<sup>464</sup> **Figure 28.** Adapté littéralement de Chaumartin (2007), p. 69 et de Steiner (2003), pp. 120-122.

### Figure 28: La conception de FrameNet en vue de ‘crime\_scenario’

Le classement thématique des unités lexicales dans FrameNet va avoir une grande influence sur la création d’un futur thésaurus, en vue de la résolution du ‘*tennis problem*’.<sup>465</sup>

#### 4.1.3 Solution du *tennis problem* : Indicateur du succès d’un thésaurus à base de *frames*

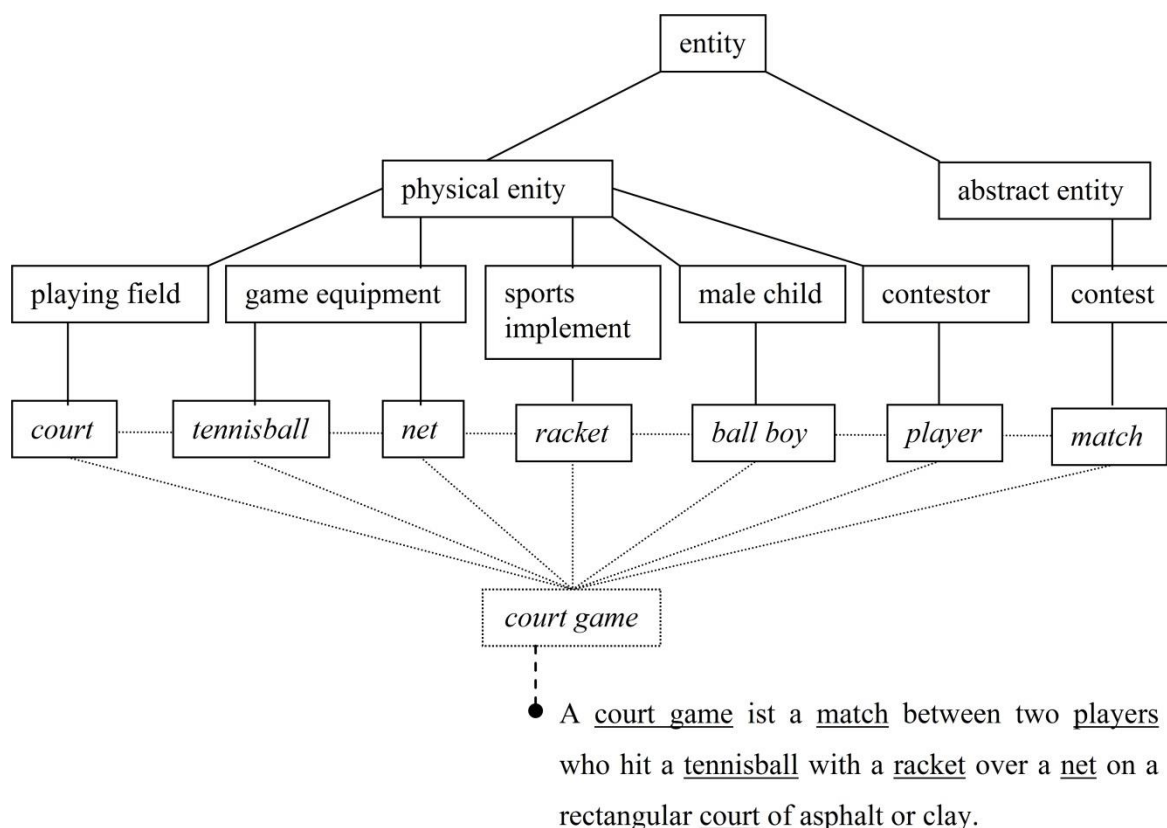
L’analyse conceptuelle des *frames* avait montré au cours de ce mémoire, que le Thésaurus Larousse présente une organisation proche de celle de FrameNet. Tout en étant organisé hiérarchiquement, comme WordNet, il présente les informations dans un grand contexte. Nous proposons, par suite, l’utilisation des relations thématiques contenues dans FrameNet et le Thésaurus Larousse pour la solution du ‘*tennis problem*’ dans WordNet. L’objectif est de développer un modèle classificatoire qui relie les différents *synsets* par attribution d’un *rôle thématique* à un cadre conceptuel plus grand, constituant un *frame*.<sup>466</sup> La figure 29 montre un extrait de la taxonomie des noms de Wordnet avec les liens thématiques qui peuvent être ajoutés pour créer un *frame* ‘court game’.

---

<sup>465</sup> Cf. Pado / Pitel (2007), p. 2.

<sup>466</sup> Cf. Approche proposé par Stevenson (2002), p. 953





\_\_\_\_ : relations classificatoires existant dans WordNet

..... : relations thématiques établies sur le Thésaurus Larousse et FrameNet

**Figure 29 : Solution du ‘tennis problem’ par ajout des liens thématiques à WordNet**  
<sup>467</sup>

Les *unités lexicales* utilisées au cours de la composition ou l’utilisation du Thésaurus Larousse évoquent des *frames*, comme celui de ‘court game’. Ainsi, un mot recherché comme ‘match’ n’évoque pas uniquement des synonymes comme ‘contest’ ou ‘competition’, mais aussi des termes comme ‘tennisball’, ‘racket’ et ‘net’. La conception du Thésaurus Larousse instaure une relation thématique entre termes sémantiquement éloignés, qui constituent un *frame* ‘court game’ au sein d’un article de dictionnaire 870. *Sports*.<sup>468</sup>

<sup>467</sup> Figure 29 de S. Reusch. Élaboré d’après un modèle proposé par Stevenson (2002), p. 953. Et, d’après les informations trouvés dans le Thésaurus Larousse (1992), Article 870. Sports, pp. 622-626 ; dans FrameNet 2.0.

<sup>468</sup> Cf. Mandala et al. (1999), p. 94. ; cf. Thésaurus Larousse (1992), article 870. Sports, pp. 622-626.

## 870 Sports

13 Tennis.- Service, service américain, service droit, service plat, service renversé. - Balle de jeu, balle de match ; ace, balle de service, balle de set. - Coup droit, drive ; demi-volée, passing-shot ; reprise de volée, revers, smash (opposé à lob), slice ; volée basse, volée haute ; amorti, brossage ou lift (opposé à chop ou coupé). - Tie-break ou jeu décisif ; double, mixte, simple.

71 Balle, ballon, palet ; batte, crosse, raquette.

79 Filet ; panier ; but, cage ; barre transversale, montants, poteaux ; bois (*les bois*) [fam.]. - Lucarne. - Ligne de but, ligne de limite, ligne de milieu, ligne de touche, surface de but. - Ligne d'arrivée, ligne de départ ; starting-block. - Plongeur, plot. - Podium. *etc.*

L'article 870. *Sports* du Thésaurus Larousse est censé former un *frame* qui décrit le dispositif normatif de 'court game'. Non seulement, il rapproche les mots recensés par Christiane Fellbaum, comme '*balle*', '*raquette*' et '*filet*', mais il présente encore des termes techniques, comme '*balle de service*', '*passing-shot*' ou '*tie-break*' qui décrivent différents scénarios de jeu. FrameNet est censé préciser les informations sur l'événement par attribution des rôles sémantiques aux *frame elements*. Ainsi, FrameNet rapproche des termes comme '*player*', '*referee*' et '*ball boy*' pour définir les acteurs du jeu (cf. figure 29).<sup>469</sup>

Le '*problème tennis*' illustre les limites de l'organisation lexicale d'un thésaurus électronique construit sur le modèle de Wordnet. Il ne peut être surmonté dans un thésaurus futur par la conception en *frames*, inspiré par l'organisation lexicale des thésauri classiques, comme le Thésaurus Larousse, et soutenu par un système électronique comme FrameNet.<sup>470</sup> Dans le chapitre prochain, nous allons préciser ce que nous attendons de la conception en *frames* en vue de la construction d'un thésaurus futur.

## 4.2 Le Thésaurus futur - Perspectives pour une gestion des connaissances

Dans le chapitre précédent nous avons traité de la comparaison du Thésaurus Larousse avec WordNet et FrameNet. Ce chapitre est la base pour une discussion sur l'élaboration d'un thésaurus futur qui va constituer le point final de cette mémoire.

---

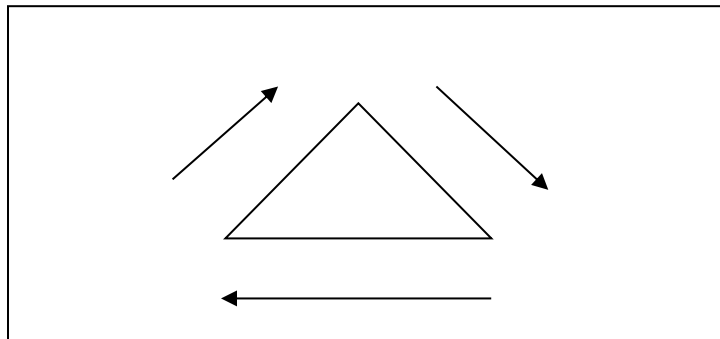
<sup>469</sup> Cf. FrameNet Version 2.0 ; cf. Fillmore / Atkins (1992), p. 131.

<sup>470</sup> Cf. Stevenson (2002), pp. 953/954 ; cf. Fillmore / Atkins (1992), pp. 123-125 ; cf. Mandala et al. (1999), pp. 97/98.

Le premier chapitre avait montré, dans un abrégé historique, qu'un thésaurus s'oriente toujours selon la mode lexicographique de son temps. On avait appris que le Dictionnaire idéologique ainsi que le Thésaurus Larousse s'inscrivaient dans la logique d'un ouvrage standard de leur époque : Roget's Thesaurus of English Words and Phrases (dans la version de 1852 et de 1987). En analogie à ces exemples, nous considérons qu'un thésaurus futur doit s'orienter également à un ouvrage standard de son temps : FrameNet, une ressource lexicale en ligne qui permet, par sa conception en '*frames*', de concevoir des applications qui facilitent une analyse de la similarité sémantique entre termes, une résolution des problèmes à l'aide de l'analogie, une spécification de la notion d'antonymie et un alignement de plusieurs ressources lexicales dans une perspective multilingue.

#### 4.2.1 Une ressource lexicale électronique constituée en *frames*

Le thésaurus futur présente une ressource lexicale électronique qui prend en compte les trois dimensions sémiotiques : l'*intension* (les propriétés de concepts), l'*extension* (les instances des concepts) et l'*expression* (les termes utilisés pour dénoter à la fois les concepts et leurs instances).<sup>471</sup>



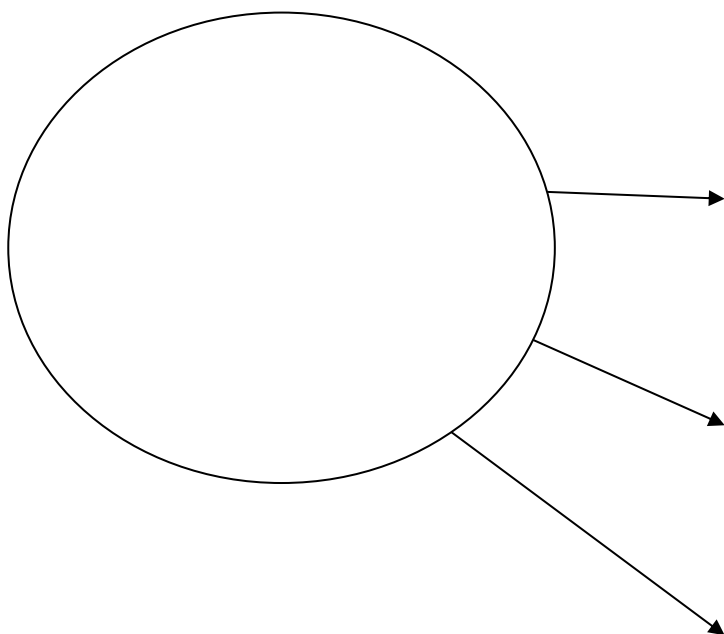
**Figure 30 : Les trois dimensions sémiotiques du futur thésaurus**<sup>472</sup>

L'utilisation d'un thésaurus s'inscrit toujours dans un *contexte* déterminé par une communauté d'usage. Le contexte d'usage n'est pas considéré dans le thésaurus classique. Dans le Thésaurus Larousse, comme dans WordNet, un *contexte formel* de classification prime sur un *contexte d'usage*. Ce *contexte formel* préfixe l'*instance attribut-objet* dans une entrée lexicale stable, de telle manière qu'elle peut être repérée par la méthode *formel concept analysis* (cf. chapitre 3.1.2). Au sein du futur thésaurus, le *contexte d'usage*

<sup>471</sup> Cf. Aimé et al. (2009), p. 229 ; Cf. Depecker (1999), pp. 42/43 ; cf. Wille (2005), p.2 ; cf. Taddei Elmi (1981), pp. 847/848 ; cf. Rey (1976), pp. 84-86.

<sup>472</sup> Figure 30. Adopté littéralement de Aimé et al. (2009), p. 229.

s'exprime par le moyen d'un *frame* qui regroupe les instances et les propriétés du concept dans une perspective d'utilisation. Le *frame*, en tant que structure d'informations dynamique, permet de réaménager spontanément les connexions entre concepts, suivant les informations recherchées.<sup>473</sup> Lors d'une étude de similarité sémantique entre termes, le programme permet d'aligner des *frames* qui ont une grande intersection. La figure 31 montre un schéma d'alignement des objets dans une perspective de similarité sémantique.



**Figure 31 : L'étude de similarité sémantique à l'aide des frames**<sup>474</sup>

Lors d'une étude d'analogie entre termes, le programme permet d'aligner des *frames* de conceptualisation éloigné, à base d'une seule intersection des ensembles (à voir une propriété d'objet). Un exemple de ce procédé sera expliqué dans le prochain sous-chapitre.

<sup>473</sup> Cf. Taylor (2009), pp. 94/95.

<sup>474</sup> Figure 31 de S.Reusch. Élaboré d'après les informations du Thésaurus Larousse (1992), 816. Transports par route, pp. 572-574 et FrameNet 2.0.

#### 4.2.2 Un auxiliaire pour le raisonnement par analogie

Le thésaurus classique est un auxiliaire de l'expression (cf. chapitre 1.2.3 et chapitre 1.3.1). Il autorise la libre exploration de l'univers des mots et des significations dans une perspective de création littéraire.<sup>475</sup> Le Thésaurus Larousse est conçu pour capturer les nuances de la langue, pour organiser et comprendre les concepts. Il dispose le vocabulaire de façon à ce que l'utilisateur y trouve tous les mots et locutions qui sont en rapport avec une idée génératrice qu'il veut rendre, groupés dans un article de dictionnaire (cf. chapitre 1.3.4). Le but d'une telle classification est que l'utilisateur, ayant une idée en tête, rencontre dans un article nécessairement une expression pour la rendre.<sup>476</sup> Dans cette perspective, une liste des mots analogues qui figure dans chaque article de dictionnaire est conçue pour offrir des aspects nouveaux et variés sur une idée *déjà en considération* ; pour agrandir la sphère de la vue intellectuelle sur un sujet *déjà en traitement*.<sup>477</sup> Le Thésaurus Larousse explicite ainsi les structures analogues uniquement à un niveau local. Une analyse globale de l'analogie ne peut être effectuée par l'utilisateur. Il manque un outil d'orientation dans la recherche des termes analogues dans le corpus. Les *références explicites* entre articles sont trop rares pour être considéré comme un instrument d'analyse fiable. Ainsi, l'utilisateur peut reconstruire un lien d'analogie entre termes *uniquement* dans un travail minutieux de lecture, page après page.

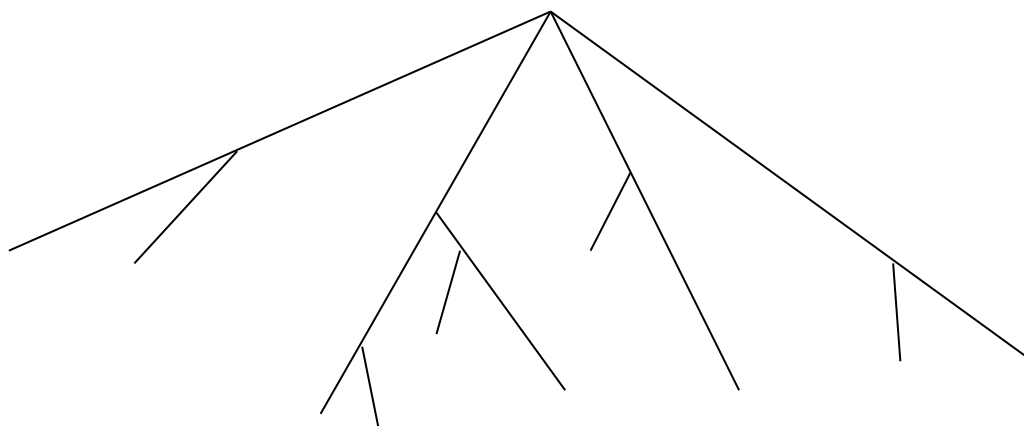
Un thésaurus devrait permettre l'analyse des liens analogiques entre termes qui appartiennent aux domaines conceptuels différents. C'est exactement ce qui serait possible dans un thésaurus futur, qui organise les concepts en '*frames*' (cf. chapitre 3.4). Dans un tel '*frame*' un concept comme '*tuyau d'eau*' est défini comme un objet possédant un diamètre, une longueur, un matériel, un épaisseur du manteau, une charge de pression maximale, un débit maximal etc. Lorsque dans un thésaurus futur, tous les concepts sont capturés dans un tel '*frame*', alors l'utilisateur pouvait consulter un attribut (par exemple, '*charge de pression maximale*') et pouvait chercher tous les objets qui comprennent cet attribut. À partir d'un premier objet ('*tuyau d'eau*'), il pouvait également rechercher tous les objets à travers lesquels coule un liquide et pouvait trouver '*boyau*' ou '*veine*'. La figure 32 montre un schéma d'alignement des objets dans une perspective d'analogie.

---

<sup>475</sup> Cf. Thésaurus Larousse (1992), Préface, p. v.

<sup>476</sup> Cf. Dictionnaire idéologique (1859), Introduction, p. v.

<sup>477</sup> Cf. Dictionnaire idéologique (1859), Introduction, p. v.



**Figure 32 : Alignement d'objet basé sur la Sémantique des cadres**<sup>478</sup>

Les '*frames*' de ces *objets cible* peuvent montrer des nouvelles propriétés qui n'ont probablement pas été pris en compte pour l'*objet source* ('*tuyau d'eau*'). Ceci peut fournir des nouvelles idées pour une modification ultérieure de cet objet.<sup>479</sup> L'alignement des objets '*tuyau*' et '*veine*' peut mener, par exemple, à de nouvelles solutions pour améliorer la circulation dans des tuyaux : comme une veine, un tuyau pourrait répondre activement aux fluctuations de pression. Un système de canalisation pourrait, dans le futur, être construit de sorte qu'un tuyau commence à pomper dès qu'il enregistre une fluctuation de pression.

La résolution des problèmes montre qu'un thésaurus futur devrait être conçu de sorte que l'utilisateur puisse atteindre avec chaque '*slot*' (attribut) de nouveaux '*frames*' (objets), pour obtenir plusieurs idées. La conception en '*frames*' part du principe que l'analogie entre termes n'est pas prédéfinie, mais dépend de la perspective de chaque utilisateur. Le thésaurus futur sera ouvert en termes d'analogie : l'utilisateur pourrait commencer à partir de chaque *objet* et à partir de chaque *attribut* une recherche des structures analogues grâce à l'assistance informatique des '*frames*' (qui reflètent la totalité des *instances attribut-*

<sup>478</sup> Figure 32 de S.Reusch. Élaboré d'après les informations du Thésaurus Larousse (1992), Articles : 331. Cœur et vaisseaux (pp. 234/235), 134. Contenant (pp. 71/72) et FrameNet 2.0.

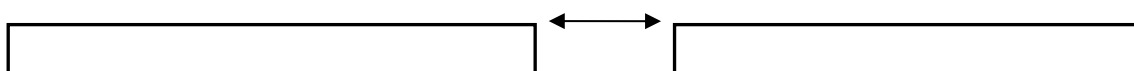
<sup>479</sup> Cf. aussi Kunzmann (1998), p.39.

*objet*). Le rôle de l'auteur en tant que médiateur des structures analogues sera levé et la porte s'ouvrira sur des expériences nouvelles pour l'utilisateur autonome.

#### 4.2.3 Un appui dans la définition des concepts par *genus* et *differentiae*

La conception du futur thésaurus en *frames* se donne pour objectif de valoriser des entités lexicales et des relations sémantiques en vue de l'univers cognitif de l'utilisateur.<sup>480</sup> Dans cette perspective, l'application des *frames* peut améliorer la spécification et l'emploi des antonymes. Dans deux occurrences d'un type de *frame* donné (qui définit le *genus*, la définition de la nature des termes), deux termes antonymes peuvent être situés, par exemple, dans des *slots* correspondants. Par leur alignement, ces *slots* établissent des *frames* « opposés ». La figure 33 montre un schéma d'alignement des *slots* dans une perspective d'antonymie.

Select : Those slots of opposite meaning & show more corresponding slots



**Figure 33 : L'étude de l'antonymie à l'aide des frames**<sup>481</sup>

Dans ces *frames* « opposés » il peut y avoir d'autres *slots* (comme '*exubérance*' et '*apathie*') qui peuvent être utilisés pour étendre la relation d'opposition entre les deux termes d'une notion supplémentaire (qui montre la *differentia*, l'écart de la notion de base par l'emploi des termes '*richesse*' et '*paucité*' pour la description des déviations comportementales).

#### 4.2.4 Une structure d'*interlingua*

Bien qu'à l'origine le thésaurus fût conçu comme un « *trésor de la langue* », il est plus qu'une représentation des connaissances linguistiques (cf. chapitre 1.1.1). Le thésaurus

<sup>480</sup> Cf. Aimé et al. (2009), p. 230.

<sup>481</sup> Figure 33 de S.Reusch. Élaboré d'après les informations du Thésaurus Larousse (1992), Articles 829. Richesse (pp. 585/586) et 830. Paucité (pp. 586/587) et de FrameNet 2.0.

peut bien être vu comme une *ontologie* qui couvre une partie substantielle des connaissances humaines. Néanmoins, ces *connaissances* sont présentées, sous forme de dictionnaire, de manière *spécifique* pour chaque langue. Les concepts et leurs relations mutuelles dans le *Thésaurus Larousse* sont déterminés, par exemple, par les modes de lexicalisation de la langue française. Ceci, parce qu'on établit la structure de l'ouvrage sur un fond des mots et des expressions inhérents au français (cf. chapitre 1.2.3). Pour toute autre langue, le *réseau sémantique* aurait une structure différente.<sup>482</sup> L'alignement des thésauri dans une perspective multilingue est ainsi loin d'être simple. Mais, il constitue un enjeu essentiel pour la lexicographie moderne.<sup>483</sup>

Le thésaurus futur présente une nouvelle approche pour construire des bases de données lexicales multilingues utilisant les *frames*.<sup>484</sup> En commençant par les informations conceptuelles contenues dans FrameNet, nous proposons un alignement avec le Thésaurus Larousse pour produire en parallèle des *frame elements* (des *unités lexicales*) en français. Les fragments résultant seront liés par le biais des *frames* de nature générique, qui fonctionnent comme une *interlingua*, une représentation linguistique neutre du contenu conceptuel partagé entre différentes langues.<sup>485</sup> Comme autrefois le *latin*, elle permet d'aligner les concepts en réduisant les irrégularités (spécificités) des différentes langues naturelles. Une fois que le *treillis* du thésaurus futur est formalisé à l'aide de *frames*, il peut aider à identifier les lacunes et les richesses lexicales de différentes langues.<sup>486</sup>

Le schéma suivant montre le fonctionnement d'une telle *interlingua*. Dans la partie inférieure du schéma, apparaissent deux *treillis* qui contiennent les mots français et anglais pour le concept '*building*'.<sup>487</sup> La grande différence entre l'anglais et le français est qu'en anglais '*house*' s'applique uniquement à des maisons d'habitation de petite taille (désignés par la lettre H, pour *house*). En français, même des bâtiments administratifs de grande taille (désignés par la lettre O, pour *office*) peuvent être appelé '*maison*'.<sup>488</sup> La Maison d'édition Larousse peut servir d'exemple.

---

<sup>482</sup> Cf. Lönneker-Rodman (2007), p. 2.

<sup>483</sup> Cf. Boas (2005), p. 445.

<sup>484</sup> Cf. Boas (2005), pp. 459-476 ; cf. Lönneker-Rodman (2007), pp.17-25 ; cf. Baker (2009), pp. 44/45.

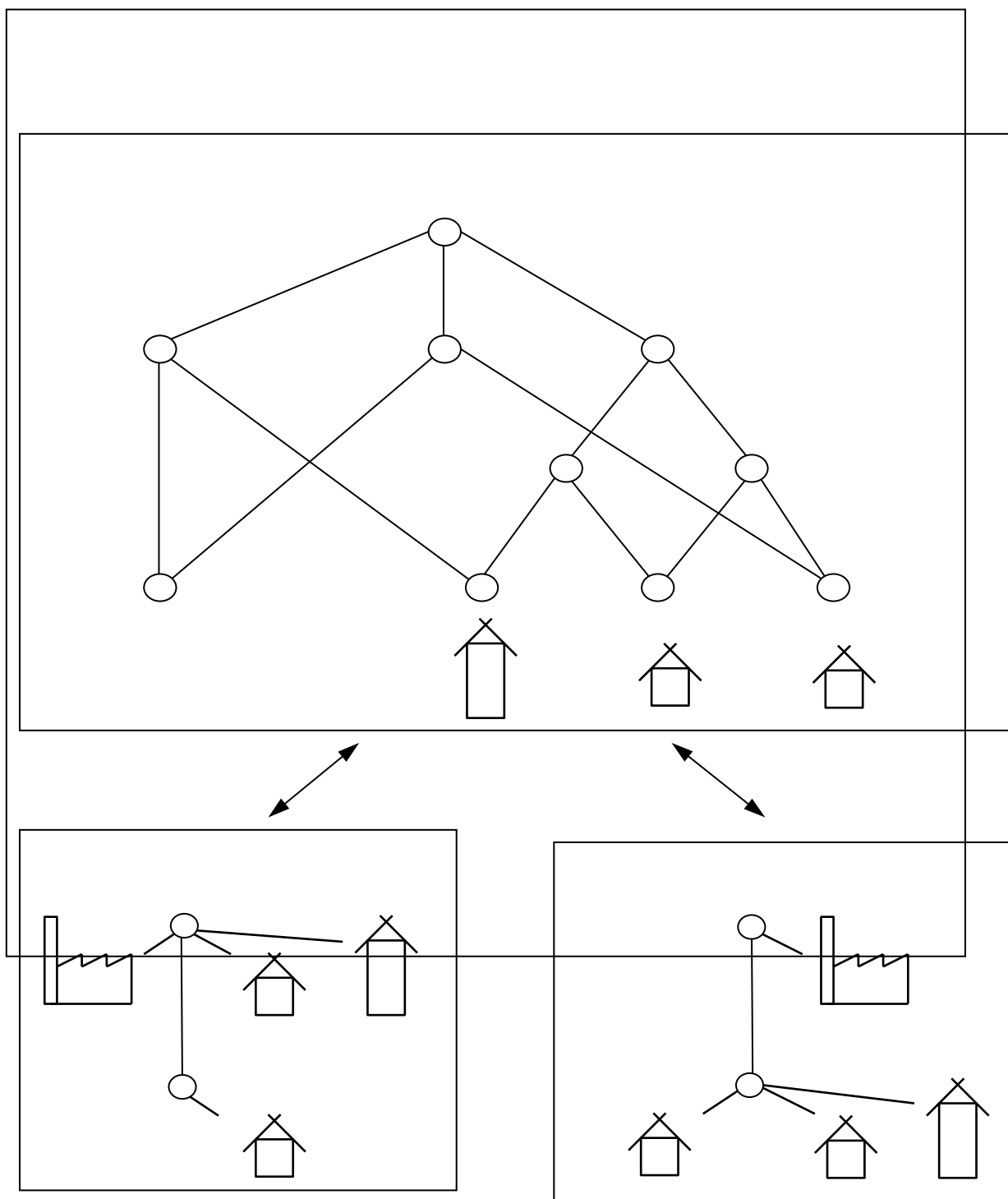
<sup>485</sup> Cf. Cardenosa et al. (2005), p. 1231 ; cf. Boas (2005), p. 445.

<sup>486</sup> Cf. Priss (2005), p.157.

<sup>487</sup> Cf. Priss (2005), pp.157/158.

<sup>488</sup> Cf. Priss (2005), pp.157/158.





**Figure 34: L'interlingua du futur thésaurus**<sup>489</sup>

<sup>489</sup> Figure 34 de S. Reusch. Élaboré d'après un schéma proposé par Priss (2005), p.159 et avec les informations tirés du Thésaurus Larousse (1992), articles 777. Architecture (pp. 535/536) et 848. Maison (pp. 600-602).

La partie supérieure du schéma présente un *treillis* qui constitue l'*interlingua* pour le thésaurus futur. Ce *treillis* fonctionne comme un '*canal d'informations*' - dans la conception de Barwise et Seligman - qui permet un alignement du *treillis* anglais et du *treillis* français.<sup>490</sup>

Pour créer un tel '*canal d'informations*', on doit identifier les *attributs* d'un *objet*, qui décrivent la différence essentielle entre l'utilisation du *mot* dans différentes langues. Il existe même déjà des approches automatisées pour des bases de données établies, tel que FrameNet.<sup>491</sup> SIMuLLDA est une telle application des bases de données multilingues, qui utilise un *treillis* de concepts comme *interlingua* en comparant automatiquement l'utilisation d'un mot dans divers langues.<sup>492</sup>

## Conclusion

Depuis 1852, le *Thesaurus of English Words and Phrases* de Peter Mark Roget est un modèle incontournable pour les créateurs des dictionnaires. En France, le *Dictionnaire idéologique* (1859) et le *Thésaurus Larousse* (1992) constituent deux approches à l'adaptation du modèle anglais. Le *Dictionnaire idéologique* - profondément imprégné par les ambitions philosophiques de vouloir déterminer les principes d'une langue universelle avec la création d'un lexique polyglotte - s'inscrit dans une « relation d'interdépendance » au *Thesaurus of English Words*. La notion d'« alignement » des deux ouvrages détermine le projet lexicographique. Théodore Robertson, le lexicographe, considère le système de classement conceptuel du *Thesaurus of English Words* comme un « moule » dans laquelle il verse le contenu lexical de la langue française - en négligeant les spécificités de la conceptualisation de chaque langue naturelle. Le *Thésaurus Larousse* - une adaptation moderne du modèle lexicographique anglais - s'inscrit dans une relation de « similarité organisationnelle », c'est-à-dire que l'ouvrage se rattache au *Thesaurus of English Words* par les mêmes principes d'organisation - sans s'approprier la conceptualisation inhérente. Le *Thésaurus Larousse* regroupe les concepts selon une vision du monde dictée par le français. Le système de classement de Peter Mark Roget sert au directeur de l'ouvrage Daniel Péchoin uniquement de « grille » pour la distribution du travail rédactionnel aux différents collaborateurs.

---

<sup>490</sup> Cf. Barwise/Seligman (1997), pp. 34 - 37 ; cf. Priss (2005), pp. 157/158.

<sup>491</sup> Cf. Baker (2009), pp. 44/45 ; cf. Priss (2005), p. 158.

<sup>492</sup> Cf. Janssen (2002), pp. 38-43 et pp. 61-107; cf. Priss (2005), p. 158.

L'analyse comparative des deux approches de l'adaptation en français du modèle lexicographique *Thesaurus of English Words* permet de préciser une démarche pour la conceptualisation d'un thésaurus futur. Le thésaurus futur doit s'orienter - comme ses prédécesseurs historiques - sur un modèle lexicographique anglais [sur une ressource électronique comme WordNet ou FrameNet] pour répondre aux exigences actuelles de la langue française, à savoir la globalisation attachée aux nouveaux médias. Une reprise des structures classificatoires proposées par un modèle anglais entraîne toujours le risque d'opter pour un classement lexical dicté par la langue source. L'« approche de dérivation » - comme le montre le cas du *Thésaurus Larousse* - atténue ce risque car il implique une évaluation critique du modèle lexicographique en vue de la conceptualisation de la langue cible. Telle est l'approche à suivre.

La création d'un futur thésaurus de la langue française est impensable sans une compréhension approfondie du « système de classement des concepts » qui constitue le fondement de ce type de dictionnaire. Un thésaurus est un ensemble ordonné des descripteurs qui constituent un système destiné à classer les concepts (par matière) et à organiser les concepts (par différentes relations sémantiques). L'analyse du *Thésaurus Larousse* unifie les deux fonctions qui sont toujours présentes dans un thésaurus, à savoir la fonction de « classification des concepts » et la fonction d'« organisation des concepts ». Le système de classification du thésaurus attribue de façon univoque et réciproque à chaque descripteur un concept. Le système d'organisation d'un thésaurus vise au recensement des relations entre concepts - comme l'antinomie, l'analogie, la similarité sémantique et la synonymie. Ces relations décident en grande partie de la classification des termes. La relation d'antonymie permet d'associer sens et contre-sens d'une notion conceptuelle - en reflétant la tendance humaine à catégoriser l'expérience en contrastes dichotomiques. Les relations de similarité sémantique, d'analogie et de synonymie permettent d'évaluer plusieurs termes en fonction d'une parenté sémantique - exprimée selon une valeur de distance que l'on peut attribuer à deux termes qui figurent dans le corpus .

Dans un premier temps, l'analyse du *Thésaurus Larousse* sert de fondement à la compréhension de la nature des termes et des relations sémantiques compris dans ce type de dictionnaire - dans la perspective de la création d'un nouveau thésaurus de la langue française.

Un thésaurus est une ressource lexicale qui repose sur des connaissances formalisées. Il capture le consensus de plusieurs rédacteurs en vue de la structuration d'un domaine de connaissance à l'aide des concepts. Sa structure - qui ressemble à un graphe mathématique - s'élabore au fur et à mesure que l'on classe les concepts dans une « hiérarchie des notions ». Elle permet d'appliquer des méthodes mathématiques (comme l'*Analyse formelle des concepts*) pour analyser le *contexte formel* d'un concept à base d'une incidence attribut-objet. Le contexte formel peut également être représenté par un *cadre* (angl. *frame*) qui permet d'étudier la structure d'informations associées à un concept plus en détail, grâce aux *elements cadres* (angl. *frame elements*) qui peuvent indiquer, par exemple, un *contenu informationnel partagé* entre deux termes similaires ou analogues. Le *cadre* est un outil qui permet en outre de relier la description sémantique aux connaissances du monde - ouvrant la perspective sur un *contexte d'usage* possible des termes figurant dans un thésaurus. La *Sémantique des cadres* est une méthode qui permet d'analyser la position du thésaurus sur le seuil entre classification lexicale et réalité linguistique. Elle montre que les termes regroupés en un domaine de connaissances - comme *printemps*, *été*, *automne* et *hiver* - ne reflètent pas uniquement un consensus des rédacteurs en vue d'une classification effectuée, mais encore une schématisation de l'expérience humaine commune aux utilisateurs d'un thésaurus.

La richesse et la particularité d'un thésaurus réside dans son organisation de connaissances par *concepts*. L'analyse du *Thésaurus Larousse* permet ainsi d'affiner la compréhension de nuances conceptuelles, par l'intermédiaire des *termes* qui s'attachent à un *cadre conceptuel* étudié. Le *cadre* - en tant que structure d'informations sous-jacente au système classificatoire au thésaurus - devient dans cette perspective l'élément fondateur sur lequel un futur thésaurus peut pendre forme.

Un futur thésaurus doit constituer une *ressource lexicale électronique* pour répondre aux exigences du monde contemporain. Dans cette perspective, l'adoption d'une structure informatique établie et communément acceptée - comme *WordNet* ou comme *FrameNet* - semble avantageuse. Une analyse contrastive avec le *Thésaurus Larousse* permet de clarifier les avantages et les limites des différentes ressources électroniques. Le logiciel de *WordNet* montre des déficiences dans la création des connexions entre différents concepts. À la différence du *Thésaurus Larousse* qui situe un concept dans un domaine de sens - *WordNet* présente à la recherche d'un descripteur (comme '*ball*')

qu'une liste des synonymes (contenant 'baseball', 'tennisball' etc.) attachés à celui-ci. Le logiciel ignore le *contexte d'usage* dans lequel peut s'inscrire un terme (comme 'ball'). Ainsi, des termes comme 'racket', 'ball' et 'net' ne sont pas censés de former un concept plus générique de 'court game'. Ces déficiences peuvent être surmontés par une ressource lexicale électronique constituée en *frames*. FrameNet est une telle ressource qui opère un « classement thématique ». Elle relie plusieurs termes (*frame elements*) à un cadre conceptuel plus grand (*frame*). Par sa conception en *frames*, la structure du logiciel se prête comme fondement pour la création d'un thésaurus futur - les connaissances acquises lors de l'analyse du Thésaurus Larousse permettant de renouveler l'approche existante par des nouvelles applications intéressantes.

Compte tenu analyses prélabes nous voyons le thésaurus futur comme une ressource lexicale électronique constituée en '*frames*' - qui s'appuie pour l'établissement de la structure informatique sur *FrameNet* et pour la conceptualisation du corpus lexicographique sur ses prédécesseurs analogues - le *Dictionnaire idéologique* et le *Thésaurus Larousse*. La constitution en '*frames*' fait du thésaurus futur un puissant moteur de recherche : sa structure dynamique permet de réaménager spontanément les liens entre concepts suivant la recherche entamée. Le *contexte formel* du logiciel et le *contexte d'usage* de l'utilisateur s'entremêlent lors de la recherche d'un terme. Un tel thésaurus n'est plus uniquement un outil d'expression, mais encore un auxiliaire pour le raisonnement par analogie. La constitution en '*frames*' permet de situer le thésaurus futur dans une perspective multilingue - sans négliger les modes de lexicalisation spécifiques des différentes langues. Une '*interlingua*' permet d'aligner les concepts en réduisant les irrégularités des langues naturelles. La réalisation de ces applications innovatrices constitue l'enjeu principal qui nous intéresse lors de la création d'un futur thésaurus de la langue française.

## Table des figures

Figure 1: Manuscrit de Peter Mark Roget .....	10
Figure 2: Le chemin de consultation identique chez Roget et Robertson .....	18
Figure 3 : Le système de classement du Thésaurus Larousse .....	34
Figure 4 : L'article 1. Existence du Thésaurus Larousse .....	35
Figure 5 : L'index du Thésaurus Larousse pour l'item "existence" .....	36
Figure 6 : Alignement symétrique des termes figurant au Thésaurus Larousse.....	38
Figure 7: Treillis exemplaire pour l'analyse formelle des concepts .....	53
Figure 8: L'analyse formelle d'un concept à l'aide des incidences attribut-objet .....	55
Figure 9 : L'analyse formelle du concept 'nombre' : Le treillis .....	58
Figure 10 : Les différentes notions de « concept » dans le Thésaurus Larousse .....	59
Figure 11: L'analyse formelle des notions d'un concept basé sur la répartition des mots .....	61
Figure 12: La similarité sémantique entre château et château fort .....	67
Figure 13: L'analogie entre tige de blé et tour de télévision .....	69
Figure 14: Métaphore créée sur la base des deux termes 'tortue' et 'enfant' .....	72
Figure 15: Le raisonnement par analogie .....	86
Figure 16: Une définition géométrique d'un rectangle .....	89
Figure 17: La hiérarchie des notions, un système qui reflète le réseau sémantique .....	93
Figure 18: Classement conceptuel du terme polysémique « avocat » dans le T.L.....	93
Figure 19: Rapprochement des termes dans l'espace sémantique d'un thésaurus .....	94
Figure 20: Application de la mesure Rada à un thésaurus .....	99
Figure 21 : Attribution des valeurs de distance aux concepts du Thésaurus Larousse .....	100
Figure 22 : Application de la mesure Resnik à un thésaurus .....	106
Figure 23: Schéma d'évaluation de distances proposées par Wu et Palmer .....	110
Figure 24: Schéma d'application de la mesure Wu/Palmer au Thésaurus Larousse .....	110
Figure 25: Calcul de similarité sémantique entre « publicité » et « produit » .....	114
Figure 26: Hypothèse soulevée par Rada - une taxonomie idéale avec des longueurs unifiées ....	117
Figure 27: Hypothèse de Wu/Palmer - une taxonomie qui se réduit à sa hiérarchie .....	119
Figure 28: La conception de FrameNet en vue de 'crime_scenario' .....	128
Figure 29 : Solution du 'tennis problem' par ajout des liens thématiques à WordNet .....	129
Figure 30 : Les trois dimensions sémiotiques du futur thésaurus .....	131
Figure 31 : L'étude de similarité sémantique à l'aide des frames .....	132
Figure 32 : Alignement d'objet basé sur la Sémantique des cadres .....	134
Figure 33 : L'étude de l'antonymie à l'aide des frames .....	135
Figure 34: L'interlingua du futur thésaurus .....	137

## Table des tableaux

Tableau 1: Comparaison de la structure classificatoire entre Roget et Robertson .....	17
Tableau 2: Divergences entre les tableaux classificatoires de Roget et Robertson .....	17
Tableau 3 : Interdépendance au niveau de la macrostructure .....	19
Tableau 4 : Interdépendance au niveau de la microstructure .....	20
Tableau 5: Les concepts redésignés par Robertson.....	21
Tableau 6 : Les articles restructurés par Robertson .....	21
Tableau 7 : Correction des arrangements critiquables par Roget .....	22
Tableau 8 : Comparaison de la structure classificatoire entre Roget et Péchoin.....	24
Tableau 9 : La transposition du « Plan of classification » au Thésaurus Larousse .....	25
Tableau 10 : Similarité organisationnelle au niveau de la macrostructure .....	26
Tableau 11 : Similarité organisationnelle au niveau de la microstructure .....	27
Tableau 12 : Le classement par descripteurs opposés dans le Thésaurus Larousse .....	37
Tableau 13 : Étude de la composition antonymique des articles.....	49
Tableau 14: Tableau exemplaire pour l'analyse formelle des concepts .....	53
Tableau 15: L'analyse formelle du concept 'nombre' : La table .....	57
Tableau 16: Étude de similarité sémantique des termes spécifiques de « maison » .....	74
Tableau 17: Étude de la relation d'analogie entre 'avion', 'montgolfière' et 'oie' .....	76
Tableau 18 : Analyse du cadre sémantique de 'château' et 'château fort' .....	80
Tableau 19 : Analyse du cadre sémantique de 'avion' et 'oie' .....	83
Tableau 20 : Un renvoi entre différentes notions d'avocat, le reflet d'un consensus adopté.....	95
Tableau 21 : : Valeurs de distance évaluées au Thésaurus Larousse .....	112

## Bibliographie

### 1. Sources

**Péchoin**, Daniel. *Thésaurus Larousse. Des idées aux mots et des mots aux idées*, Paris 1992.

**Péchoin**, Daniel. *Dictionnaire des analogies*, Paris 2009.

**Robertson**, Théodore. *Dictionnaire idéologique. Recueil des mots, des phrases, des idiotismes et des proverbes de la langue française classés selon l'ordre des idées*, Paris - Londres 1859.

**Roget**, Peter Mark. *Thesaurus of English words and Phrases*. 150<sup>th</sup> Anniversary Edition edited by George Davidson. Londres 2002.

**Roget**, Peter Mark. *Thesaurus of English words and Phrases*. Londres 1856. Consultable sur : <http://books.google.com/books?id=9nYCAAAAQAAJ&hl=de>

**Roget**, Peter Mark. *Thesaurus of English words and Phrases*. Londres 1911. Téléchargeable sur : <http://www.gutenberg.org/ebooks/10681>

**Roget**, Peter Mark. *Thesaurus of English words and Phrases*. Harlow -Essex 1987.

**WordNet 2.1**. Ressource lexicale électronique de l'Université de Princeton, consultable sur : <http://wordnet.princeton.edu/> .

**FrameNet II**. Ressource lexicale électronique de l'Université de Berkeley, consultable sur : <https://framenet.icsi.berkeley.edu>

### 2. Dictionnaires

**Boissière**, Prudence. *Dictionnaire analogique de la langue française : répertoire complet des mots par les idées et les idées par les mots*. Paris 1862.

**Dubois**, Jean / Giacomo, Mathée / Guespin, Louis, Marcellesi, Christiane / Marcellesi, Jean-Baptiste / Mével, Jean-Pierre, *Dictionnaire de linguistique*, Paris 1994.



**Maquet**, Charles. *Dictionnaire analogique. Répertoire moderne des mots par les idées, des idées par les mots d'après les principes de P. Boissière*, Paris 1936.

**Niobey**, Georges. *Dictionnaire analogique*. Paris 2007.

**Rey-Debove**, Josette / **Rey**, Alain. *Le Nouveau Petit Robert . Dictionnaire alphabétique et analogique de la langue française*. Paris 2007.

### 3. Ouvrages

**Barwise**, Kenneth Jon / **Seligman**, Jerry. *Information Flow: the Logic of Distributed Systems*, Cambridge 1997.

**Caspar**, Nathalie / **Leclerc**, Bruno / **Monjardet**, Bernard. *Ensembles ordonnés finis : concepts, résultats et usages*. Berlin - Heidelberg - New York 2007.

**Cassirer**, Ernst : *Substance et fonction. Éléments pour une théorie du concept*, Paris 1977.

**Charconnet**, Jean. *Analogie et logique naturelle. Une étude des traces linguistiques du raisonnement analogique à travers différents discours*. Berne 2003.

**Cogis**, Oliver/ **Robert**, Claudine. *Au-delà des ponts de Königsberg. Théorie des graphes. Problèmes, théorèmes, algorithmes*. Paris 2003.

**Depecker**, Loïs. *Entre signe et concept. Éléments pour une terminologie générale*. Presses de la Sorbonne nouvelle, Paris 2002.

**Fellbaum**, Christiane (Éd.). *WordNet: an Electronic Lexical Database*, MIT Press, Cambridge 1998.

**Hüllen**, Werner, *Networks and Knowledge in Roget's Thesaurus. From ancient to medieval*, Oxford 2009.

**Hüllen**, Werner, *A History of Roget's Thesaurus: Origins, Development, and Design*. Oxford: Oxford University Press. 2005.

**Itkonen**, Esa: *Analogy as structure and process. Approches in linguistics, cognitive psychology and philosophy of science*. Philadelphia. Amsterdam 2005.

**Koester**, Bjoern. *FooCA. Web information retrieval with formal concept analysis. Beiträge zur begrifflichen Wissensverarbeitung (Contributions to Conceptual Knowledge Processing)*, Verlag allgemeine Wissenschaft - HRW e.K., Mühlthal, 2006.

**Konerdin**, Klaus-Peter / **Henne**, Helmut / **Sitta**, Horst / **Wiegand**, Herbert Ernst (Éds.): *Frames und lexikalisches Bedeutungswissen. Untersuchungen zur linguistischen Grundlegung einer Frametheorie und zu ihrer Anwendung in der Lexikographie*. Max Niemeyer Verlag, Tübingen 1993.

**Kunzmann, Peter.** *Dimensionen von Analogie. Wittgensteins Neuentdeckung eines klassischen Prinzips.* Düsseldorf - Bonn 1998.

**Lehmann, Alise / Martin-Berthet, Françoise.** *Introduction à la lexicologie. Sémantique et morphologie.* Nathan, Paris 2000 (1998<sup>1</sup>).

**Matoré, Georges.** *Histoire des dictionnaires français,* Paris 1968.

**Mel' čuk, Igor / Clas, André / Polguère, Alain.** *Introduction à la lexicologie explicative et combinatoire,* Louvain-la Neuve 1995.

**Martinet, André.** *Éléments de linguistique générale,* Armand Colin, Paris 1967.

**Monneret, Phillipe.** *Essais de linguistique analogique,* Association Bourguignonne d'Études Linguistiques et Littéraires. Dijon 2004.

**Niklas-Salminen, Aïno.** *La lexicologie.* Paris 1997.

**Polguère, Alain.** *Lexicologie et sémantique lexicale. Notions fondamentales.* Presses de l'université de Montréal, Montréal 2003.

**Rey, Alain.** *Théories du signe et du sens, Tome I. Lectures,* Paris 1973.

**Rey, Alain.** *Théories du signe et du sens, Tome II. Lectures,* Paris 1976.

**Rey-Debove, Josette.** *Étude linguistique et sémiotique des dictionnaires français contemporains,* Paris 1971.

**Reisinger, Leo.** *Rechtsinformatik,* W. de Gruyter, Berlin 1977.

**Reusch, Peter Johann Adam (1988),** *Modellverwaltung und Expertensystemkomponenten für betriebliche Informationssysteme,* BI Wissenschaftsverlag, Mannheim - Wien - Zürich, 1988.

**Sander, Emmanuel.** *L'analogie du Naïf au créatif. Analogie et catégorisation,* Paris 2000.

**Stuckenschmidt, Heiner.** *Ontologien: Konzepte, Technologien und Anwendungen.* Springer Verlag. Berlin - Heidelberg 2009.

**Tijus, Charles.** *Métaphores et analogies,* Paris 2003.

**Quemada, Bernard.** *Les Dictionnaires du français moderne (1539-1863). Étude sur leur histoire, leurs types et leurs méthodes,* Paris 1967.

**Wille, Rudolf / Ganter, Bernhard.** *Formale Begriffsanalyse. Mathematische Grundlagen,* Springer Verlag, Berlin. Heidelberg, 1996.

**Zemmour, David.** *Initiation à la linguistique.* Paris 2008.



#### 4. Mémoires de diplôme

**Schwab**, Didier. *Vecteurs conceptuels et fonctions lexicales : Application à l'antonymie*. Diplôme d'études approfondies en Informatique de l'Université de Montpellier II. Sciences et Techniques du Languedoc, Montpellier 2001. Consultable sur :

[http://atoum.imag.fr/geta/User/didier.schwab/Publications/memoire\\_DEA\\_Didier\\_SCHWAB.pdf](http://atoum.imag.fr/geta/User/didier.schwab/Publications/memoire_DEA_Didier_SCHWAB.pdf)

#### 5. Thèses de doctorat

**Archer**, Vincent. *Graphes linguistiques multiniveau pour l'extraction de connaissances : l'exemple des collocations*, Thèse de doctorat, L'UNIVERSITÉ JOSEPH FOURIER Grenoble 1, September 2009 Grenoble. Consultable sur :

<http://tel.archives-ouvertes.fr/docs/00/42/65/17/PDF/Archer-These.pdf>

**Batteux**, Martina. *Die französische Synonymie im Spannungsfeld zwischen Syntagmatik und Paradigmatik*. Publié dans: Publications Universitaires Européennes, Série XIII. Langue et littératures françaises, Volume 255, Francfort - Berlin - Berne - Bruxelles - New York - Oxford - Vienne 2000. Consultable sur :

<http://edoc.hu-berlin.de/dissertationen/philologie/batteux-martina/PDF/Batteux.pdf>

**Jannsen**, Maarten. *SIMuLLDA. A Multilingual Lexical Database Application using a Structured Interlingua*. PhD Thesis, Universiteit Utrecht, Utrecht 2002. Consultable sur:

<http://maarten.janssenweb.net/simullda/proefschrift.pdf>

**Jarmasz**, Mario. *Roget's thesaurus as a lexical resource for natural language processing*, Thesis submitted to the faculty of graduate and postdoctoral studies for the degree of Master of Computer Science, Ottawa-Carleton Institute for Computer Science, University of Ottawa, Ontario July 2003. Consultable sur:

[http://www.site.uottawa.ca/~mjarmasz/thesis/Roget\\_for\\_NLP.pdf](http://www.site.uottawa.ca/~mjarmasz/thesis/Roget_for_NLP.pdf)

**Mazuel**, Laurent. *Traitement de l'hétérogénéité sémantique dans les interactions humain-agent et agent-agent*. Thèse de doctorat de Université Paris VI Pierre & Marie Curie, soutenue le 21 novembre 2008, Paris 2008. Consultable sur :

<http://pertomed.spim.jussieu.fr/~lma/publis/pdf/These-LaurentMazuel-v20090127.pdf>

**Old**, John. *The semantic structure of Roget's. A whole language thesaurus*. Indiana, 2003. Consultable sur :

<http://www.johnold.org/LJOLD/papers/Dissertation.pdf>

**Schwab**, Didier. *Approche hybride - lexicale et thématique - pour la modélisation, la détection et l'exploitation des fonctions lexicales en vue de l'analyse sémantique de texte*, Thèse de Doctorat de l'Informatique l'Université de Montpellier II. Sciences et Techniques du Languedoc, soutenu le 7 décembre 2005, Montpellier 2005. Consultable sur :

[http://www.lirmm.fr/~schwab/Publications/these\\_Schwab\\_version\\_ecran.pdf](http://www.lirmm.fr/~schwab/Publications/these_Schwab_version_ecran.pdf)

## 6. Publications

**Agirre**, Eneko / Rigau, German (1996). "Word Sense Disambiguation Using Conceptual Density", In: *Proceedings of the 16th International Conference on Computational Linguistics* (Coling '96), Copenhagen 1996, pp.16-22. Consultable sur: <http://dl.acm.org/citation.cfm?id=992635>

**Aimé**, Xavier / Fürst, Frédéric / Kuntz, Pascale / Trichet, Francky (2009). "SEMIOSEM : une mesure de similarité conceptuelle fondée sur une approche sémiotique". In : Fabien L. Gandon (ed.). *Actes des 20es Journées Francophones d'Ingénierie des Connaissances*, Hammamet (Tunisie), 25.-29. Mai 2009, pp. 229-240. Consultable sur : [http://hal.archives-ouvertes.fr/hal-00377546\\_v1/](http://hal.archives-ouvertes.fr/hal-00377546_v1/)

**Baker**, Collin F. (2009). "La sémantique des cadres et le projet FRAMENET : une approche différente de la notion de « valence »". In : *Langages* 4/2009 (n° 176), pp. 32-49.

**Bianucci**, Gabriele (1992). "L'usage de relations sémantiques dans l'élaboration de thésauri : L'expérience du PTP (Petit Thesaurus Politique)". In : *Cahiers de lexicologie. Revue internationale de lexicologie et lexicographie*. 1992-2, pp. 59-84.

**Bloch**, Oscar (1904). "Étude sur le Dictionnaire de J. Nicot (1606)", In : *Mélanges de philologie offerts à F. Brunot*, Société Nouvelle de Librairie et d'Édition, Paris 1904, pp. 1-13.

**Boas**, Hans C. (2005). "Semantic Frames as Interlingual Representations for Multilingual Lexical Databases". In: *International Journal of Lexicography*, Vol. 18 (4), Decembre 2005, pp. 445-478.

**Boulanger**, Jean-Claude (1993). "PÉCHOIN, Daniel (dir.) (1991) : Thésaurus Larousse. Des mots aux idées, des idées aux mots , Paris, Larousse, XXI + 146 p". In : *Meta : Journal des traducteurs/Meta : Translators' journal*, vol. 38, n°1, mars 1993, pp.130-136. Consultable sur : <http://www.erudit.org/revue/meta/1993/v38/n1/002120ar.pdf>

**Cardeñosa**, Jésus / Gallardo, Carloina / Iraola, Luis (2005). "Using an Interlingua for Document Knowledge Representation". In: *Proceedings of 4<sup>th</sup> Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)*, 9 Septembre 2005 Barcelona, pp. 1231-1236.

**Chaumartin**, François-Régis (2007). "WordNet et son écosystème : constitution et utilisation de ressources linguistiques de large couverture". In : Nam, Jessun / Polguère, Alain. *Procédés de BDL-CA - Bases de données lexicales : construction et applications. Colloque organisé par l'Observatoire de linguistique Sens-Texte (OLST)*, Montréal 23 avril 2007, pp. 59-79. Consultable sur :

<http://olst.ling.umontreal.ca/pdf/BDL-CA07.pdf>

**De Tollenaere**, Félicien (1960). "Lexicographie alphabétique ou idéologique". In: *Cahiers de lexicologie. Revue internationale de lexicologie et lexicographie*, 1960-2, pp.19-29.

**Fellbaum**, Christiane (1999). "La représentation des verbes dans le réseau sémantique WordNet". In: *Langages*, 33<sup>e</sup> année, n°136, 1999. pp. 27-40. Consultable sur:  
[http://www.persee.fr/web/revues/home/prescript/article/lgge\\_0458-726x\\_1999\\_num\\_33\\_136\\_2210](http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726x_1999_num_33_136_2210)

**Fillmore**, Charles J. (1975). "An alternative to Checklist Theories of Meaning", In: *Proceedings of the First Annual Meetings of the Berkeley Linguistics Society (BLS)*, Berkeley, California, pp. 123-131, February 1975, Consultable sur :

<http://www.icsi.berkeley.edu/cgi-bin/pubs/publication.pl?ID=2868>

**Fillmore**, Charles J. (1985). "Frames and the Semantics of Understanding", In: *Quaderni di Semantica*, Vol. 6, N° 2, Bologna, Il Mulino, pp. 222-254. Consultable sur:

<http://www.icsi.berkeley.edu/cgi-bin/pubs/publication.pl?ID=1847>

**Fillmore**, Charles J. / Atkins, Beryl T. (1992). "Toward a Frame-Based Lexicon: The Semantics of RISK and its Neighbors." In: Adrienne Leherer et Eva Feder Kittay (Eds.), *Frames, Fields, and Contrasts : New Essays in Semantic and Lexical Organization*, Hillsdale, N.J.: Laurence Erlbaum Associates, pp.75-102. Consultable sur:

<http://www9.georgetown.edu/faculty/ard8/Ling565/1992.pdf>

**Gandon**, Fabien / Corby, Olivier / Diop, Ibrahima / Lo, Moussa (2008). "Distances sémantiques dans des applications de gestion d'information utilisant le web sémantique." In : *8<sup>èmes</sup> Journées Francophones. Extraction et Gestion des Connaissances. Mesures de similarité sémantique*. Sophia 29. Janvier 2008, pp. 87-96.

**Harrathi**, Rami / Calabretto, Sylvie (2010). "Une approche de recherche sémantique dans les documents sémi-structurés". In : *7<sup>ième</sup> Conférence en Recherche d'Informations et Applications - CORIA 2010*, Sousse 18.-20. Mars 2010, 20 pages, Consultable sur :

[http://liris.cnrs.fr/~sbc/sbcwiki/images/c/c2/RISE2010\\_2.pdf](http://liris.cnrs.fr/~sbc/sbcwiki/images/c/c2/RISE2010_2.pdf)

**Hodgson, Nathalie** (1997). "Analyse documentaire III : Thesaurus et fichier d'autorité à l'Université Concordia", In : *Archives*, Volume 28-3/4 (1996-1997), pp. 43-56. Consultable sur :

[http://www.archivistes.qc.ca/revuearchives/vol28\\_3\\_4/28-3&4-Hodgson.pdf](http://www.archivistes.qc.ca/revuearchives/vol28_3_4/28-3&4-Hodgson.pdf)

**Holyoak, Keith J. / Koh, Kyunghee** (1987). "Surface and structural similarity in analogical transfer". In: *Memory and Cognition*, 15, pp. 332-340.

**Jarmasz, Mario / Szpakowicz, Stan** (2003). "Roget's Thesaurus and Semantic Similarity". In: *Proceedings of Conference on Recent Advances in Natural Language Processing (RANLP 2003)*, Borovets September 2003, pp. 212-219. Consultable sur:

[citeseerx.ist.psu.edu](http://citeseerx.ist.psu.edu)

**Khatraoui, Mouna / Bousbia, Nabila / Balla, Amar** (2008). "Détection de la similarité sémantique entre pages visitées durant une session d'apprentissage". In : Aufaure, Marie-Aude / Boussaid, Ormar / Kuntz, Pascale (Eds.). *8<sup>ième</sup> Journées Francophones sur l'Extraction et Gestion des Connaissances. Mesures de similarité sémantique (EGC 2008)*, Sophia 29. janvier 2008, pp. 121-129. Consultable sur :

[www.egc.asso.fr/sdoc-94-egc08\\_atelier\\_mss.pdf](http://www.egc.asso.fr/sdoc-94-egc08_atelier_mss.pdf)

**Lafourcade, Mathieu / Prince, Violaine** (2001). "Synonymie et vecteurs conceptuels". In: *Proceedings of TALN 2001*, Tours 2.-5. juillet 2001, pp. 233 - 242. Consultable sur : <http://www.lirmm.fr/~lafourca/ML-biblio/TALN2001/ML-taln2001-vf.pdf>

**Lamizet, Bernard** (1975). "Pour une approche formelle de la synonymie lexicographique". In : *Cahiers de lexicologie. Revue internationale de lexicologie et lexicographie*, 1975-26, pp. 15-46.

**Lönneker-Rodman, Birte** (2007). *Multilinguality and FrameNet*, International Computer Science Institute Berkeley California, Technic Report TR-07-001, Mars 2007, 25 pages. Consultable sur:

<http://www.icsi.berkeley.edu/pubs/techreports/tr-07-001.pdf>

**Maala, Mohamed Zied / Delteil, Alexandre / Napoli, Amedeo** (2007). "Distance sémantique entre concepts définis en ALE". In : *Conférence Francophone Langues et modèles à objets (LMO 2007)*, pp. 117-130. Consultable sur :

<http://hal.inria.fr/docs/00/20/15/68/PDF/an-lmo-150107.pdf>

**Mandala, Rila / Tokunaga, Takenobu / Tanaka, Hozumi** (1999). "Complementing WordNet with Roget's and Corpus-based Thesauri for Information Retrieval". In : *Proceedings of the 9<sup>th</sup> Conference of the European Chapter of the Association for*

*Computational Linguistics (EACL 1999)*, Bergen 8.-12. Juin 1999, pp. 94-101. Consultable sur:

<http://acl.ldc.upenn.edu/E/E99/E99-1013.pdf>

**Mansour**, Hormoz (1983), "A Structural Approach to Analogy", In: *A.I. Memo*, N° 747, Massachusetts Institute of Technologie, Artificial Intelligence Laboratory, Novembre 1983, 27 pages . Consultable sur :

<http://dspace.mit.edu/handle/1721.1/5657>

**Marello**, Carla. "The Thesaurus". In : Hausmann, Franz Josef/ Reichmann, Oskar, Wiegand, Herbert Ernst/ Zgusta, Ladislav. *Encyclopédie internationale de lexicographie*, Tome I, Berlin - New York 1989, pp. 1083-1093.

**Mc Hale**, Michael (1998). "A Comparison of WordNet and Roget' s Taxonomy for Measuring Semantic Similarity". In: *Proceedings of COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems*. Montreal August 1998, pp.115-120. Consultable sur :

<acl.ldc.upenn.edu/W/W98/W98-0716.pdf>

**Minsky**, Marvin Lee (1975). "A Framework for Representing Knowledge". In: Patrick Henry Winston: *The Psychology of Computer Vision*, New York, pp. 211-278. Consultable sur :

<http://dspace.mit.edu/handle/1721.1/6089>

**Minsky**, Marvin Lee (1980). "Jokes and their Relation to the Cognitive Unconscious". In: *A.I. Memo N° 603*, Massachusetts Institute of Technologie, Artificial Intelligence Laboratory, Massachusetts Novembre 1980, 25 pages. Consultable sur:

<http://dspace.mit.edu/handle/1721.1/5701> et sur

<http://web.media.mit.edu/~minsky/papers/jokes.cognitive.txt>

**Oim**, Haldur/ Saluveer, Madis (1985). "Frames in Linguistic Descriptions". In : *Quaderni di Semantica*, Vol. 6, N° 2., Bologna, Il Mulino, pp. 295-305.

**Old**, John L. / Priss, Uta (2001). "Metaphor and Information Flow". In: *Proceedings of the 12th Midwest Artificial Intelligence and Cognitive Science Conference*, pp. 99-104. Consultable sur: <http://www.upriss.org.uk/top/research.html>

**Old**, John L. (2004). "Unlocking the Semantics of Roget's Thesaurus". In: P. Eklund, (Ed.), *2nd International Conference on Formal Concept Analysis, Lecture Notes in Computer Science*, Vol. 2961, pp. 236-243. Consultable sur:

<http://www.johnold.org/LJOLD/papers/Roget6.pdf>



**Padó**, Sebastian / Pitel, Guillaume (2007). “Annotation précise du français en sémantique de rôles par projection cross-linguistique”. In : *Proceedings of TALN 2007*, Toulouse, 12-15 juin 2007, 6 pages. Consultable sur :

[http://www.nlpado.de/~sebastian/pub/papers/taln07\\_pado.pdf](http://www.nlpado.de/~sebastian/pub/papers/taln07_pado.pdf)

**Péchoin**, Daniel (2001). “Thésaurus documentaire, thésaurus idéologique et dictionnaire analogique”. In : J.-P. Cotten / B. Hufschmitt. *Repérer, formaliser, traduire les concepts philosophiques*. Colloque de Besancon 23-24 juin 1999, Paris 2001.

**Petruck**, Miriam (1996). “Frame Semantics”. In: Jef Verschueren, Jan-Ola Östman, Jan Blommaert, und Chris Bulcaen (Eds.): *Handbook of Pragmatics*. John Benjamins, Amsterdam & Philadelphia 1996, pp. 381-399. Consultable sur internet :

<http://www.google.de/url?sa=t&source=web&cd=1&ved=0CBwQFjAA&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.4.7422%26rep%3Drep1%26type%3Dpdf&rct=j&q=Miriam%20R.%20L.%20Petruck%20frame%20semantics&ei=lWCATqEr8LTRBZvT6LkJ&usg=AFQjCNE9Z3UO-g9LZvbbj3iyVqaYGtXFLA>

**Priss**, Uta (2005). “Linguistic Applications of Formal Concept Analysis”. In: Bernhard Ganter, Gerd Stumme, Rudolf Wille (eds.): *Formal Concept Analysis. Foundations and Applications*, Springer, Berlin 2005, pp. 149-160. Consultable sur :

[www. google.books.de](http://www.google.books.de) [par recherche du titre de la publication]

**Rada**, Roy / Hafedh Mili / Ellen Bicknell / Maria Blettner (1989), “Development and Application of a metric on Semantic Nets”, In: *IEEE Transaction on Systems, Man and cybernetics*, vol. 19(1), pp. 17-30. Consultable sur:

[http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=24528](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=24528)

**Renard** Arnaud, Sylvie Calabretto, Béatrice Rumpler, (2009) “Recherche d’Information Sémantique: Appariement sémantique flou de documents semi-structurés”, In: C. Roussey et J. P. Chevallet (Eds.), *27<sup>ème</sup> Congrès INFORSID, Atelier Recherche d'Information Sémantique RISE*, Toulouse 2009, pp. 86-102. Consultable sur :

<http://liris.cnrs.fr/~sbc/sbcwiki/images/b/ba/RISE2009Renard.pdf>

**Resnik**, Philip (1995). “Using Information Content to Evaluate Semantic Similarity in a Taxonomy”. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal 20. - 25. Août 1995, pp. 448–453. Consultable sur:

<http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.55.5277>

**Resnik**, Philip (1999). “Semantic similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language”. In: *Journal of Artificial Intelligence Research No. 11*, pp. 95-130. Consultable sur:

<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.3785>

**Reusch**, Sylvie (2011a). "Project Thesaurus 2020." In: *Current Issues in Management of Business and Society Development 2011*. University of Latvia, Riga 5. - 7. May 2011.

**Reusch**, Sylvie (2011b). "Project Thesaurus 2020 - Linguistic and Ontological Aspects". In: *Proceedings of the 6<sup>th</sup> IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, Prague 15.-17. Septembre 2011, pp. 937 - 941.

**Rey-Debove**, Josette (1989). "Les systèmes de renvois dans le dictionnaire monolingue". In : Hausmann, Franz Josef/ Reichmann, Oskar, Wiegand, Herbert Ernst/ Zgusta, Ladislav. *Encyclopédie internationale de lexicographie*, Tome I, Berlin - New York 1989, pp. 931-937.

**Safar**, Brigitte / Reynaud, Chantal / Calvier, François (2007). "Techniques d'alignement d'ontologies basées sur la structure d'une ressource complémentaire". In : *Journées Francophones sur les Ontologies*, JFO-07, Sousse 18-20 Octobre 2007, pp. 21-35.

**Schwab**, Didier / Lafourcade, Mathieu / Prince, Violaine (2002a). "Vers l'apprentissage automatique pour et par les vecteurs conceptuels de fonctions lexicales — l'exemple de l'antonymie". In : *Proceedings of TALN 2002*, Nancy 24.-27. Juin 2002, pp. 125-134. Consultable sur :

[http://www.atala.org/doc/actes\\_taln/AC\\_0070.pdf](http://www.atala.org/doc/actes_taln/AC_0070.pdf)

**Schwab**, Didier / Lafourcade, Mathieu / Prince, Violaine (2002b). "Amélioration de la représentation sémantique lexicale par des vecteurs conceptuels : le rôle de l'antonymie". In : *Actes des 6<sup>èmes</sup> journées internationales d'analyse statistique des données textuelles (JADT 2002)*. Saint-Malo 13. - 15. Mars 2002, pp. 701-712. Consultable sur :

[http://lexicometrica.univ-paris3.fr/jadt/jadt2002/PDF-2002/schwab\\_lafourcade\\_prince.pdf](http://lexicometrica.univ-paris3.fr/jadt/jadt2002/PDF-2002/schwab_lafourcade_prince.pdf)

**Schwab**, Didier / Lafourcade, Mathieu / Prince, Violaine (2003). "Amélioration de liens entre acceptions par fonctions lexicales vectorielles symétriques". In: *Proceedings of TALN 2003*, Batz-sur-Mer 11. - 14. juin 2001, pp. 235-244. Consultable sur :

[http://www.atala.org/doc/actes\\_taln/AC\\_0098.pdf](http://www.atala.org/doc/actes_taln/AC_0098.pdf)

**Schwab**, Didier / Lafourcade, Mathieu / Prince, Violaine (2005). "Extraction semi-supervisée de couples d'antonymes grâce à leur morphologie". In : *Proceedings of TALN 2005*, Dourdan 6. - 10. juin 2005, pp. 73 - 82. Consultable sur :

[http://atoum.imag.fr/geta/User/didier.schwab/Publications/Schwab\\_Taln2005FR.pdf](http://atoum.imag.fr/geta/User/didier.schwab/Publications/Schwab_Taln2005FR.pdf)

**Sedelow**, Sally Y. (1993). "The formal analysis of concepts", in: *Behavior Research Methods, Instruments & Computers*, Volume 25 (2), 1993, pp. 314-317. Consultable sur: <http://www.springerlink.com/content/k8kw477628842096/fulltext.pdf>

**Slimani**, Thabet / Yaghlane, Boutheina Ben / Mellouli, Khaled (2007). "Une extension de mesure de similarité entre les concepts d'une ontologie". In : *4th International Conference:*

*Sciences of Electronic, Technologies of Information and Telecommunications*, Setit 25.-29. Mars 2007, 10 pages. Consultable sur :

[http://www.setit.rnu.tn/last\\_edition/setit2007/TI/86.pdf](http://www.setit.rnu.tn/last_edition/setit2007/TI/86.pdf)

**Steiner**, Petra (2003). "FrameNet und WordNet: Perspektiven für die Verknüpfung zweier lexikalisch-semantischer Netze". In: Kunze, Claudia / Lemnitzer, Lothar / Wagner, Andreas (Eds.). *GermaNet-Workshop: Anwendungen des deutschen Wortnetzes in Theorie und Praxis. Tagungsband*. Tübingen: Universität Tübingen, Seminar für Sprachwissenschaft, Abt. Computerlinguistik, 2003. pp. 119-138. Consultable sur:

<http://arbuckle.sfs.uni-tuebingen.de/GermaNet/GermaNet-Workshop/PSteiner.pdf>

**Stevenson**, Mark (2002). "Augmenting Noun Taxonomies by Combining Lexical Similarity Metrics". In: *Proceedings of the 19th International Conference on Computational Linguistics (COLING-02)*, Taipei 24. Août - 1. Septembre 2002, pp. 953-959. Consultable sur:

<http://aclweb.org/anthology/C/C02/C02-1038.pdf>

**Taddei Elmi**, Giancarlo (1981). "Synonymie, hyponymie et thésaurus juridiques", dans : *Actes du Congrès international informatique et sciences humaines*, Université de Liège, Liège 1981, pp. 845-852. Consultable sur :

<http://promethee.philo.ulg.ac.be/LASLApdf/Actes/78.pdf>

**Wagner**, Hans: "Begriff". In: H. Krings, H.M. Baumgartner, C. Wild (Eds.). *Handbuch philosophischer Grundbegriffe*. Kösel, München 1973, pp. 191-209.

**Wille**, Rudolf (1987). "Bedeutungen von Begriffsverbänden". In: Bernhard Ganter / Rudolf Wille / Karl-Erich Wolff (Eds.), *Beiträge zur Begriffsanalyse*, BI Wissenschaftsverlag, München - Wien - Zürich, 1987.

**Wille**, Rudolf (2005). „Formal Concept Analysis as Mathematical Theory“. In: Bernhard Ganter, Gerd Stumme, Rudolf Wille (eds.): *Formal Concept Analysis. Foundations and Applications*, Springer, Berlin 2005, pp. 1-33. Consultable sur :

[www. google.books.de](http://www.google.books.de), par recherche du titre de la publication.

**Wu**, Zhibiao / Palmer Martha (1994). "Verb Semantics and Lexical Selection". In: *Proceedings of the 32nd Annual meetings of the Associations for Computational Linguistics*, New Mexico State University, Las Cruces 27.-30. Juin 1994, pp. 133-138. Consultable sur:

<http://acl.ldc.upenn.edu/P/P94/P94-1019.pdf>

**Veale**, Tony (2006). "A typology of Lexical Analogy in WordNet". In: *Proceedings of the 3rd Global WordNet Conference*, Seogwipo (Korea) 22. - 26. Janvier 2006, pp. 105-110. Consultable sur:

<http://afflatus.ucd.ie/papers/GWC2006b.pdf>

**Ventresque**, Anthony (2006). “Une mesure de similarité sémantique utilisant des résultats de psychologie”. In : *Proceedings of CORIA 2006. Conférence en Recherche d'Informations et Applications*, Lyon 15.-17.Mars 2006, pp. 371-376. Consultable sur : <http://hal.univ-nantes.fr/docs/00/41/96/14/PDF/371.pdf>

**Ventresque**, Anthony / Cazalens, Sylvie / Lamarre, Philippe / Valduriez / Patrick (2008). “Enrichissement sémantique de requêtes utilisant un ordre sur les concepts”, In : Aufaure, Marie-Aude / Boussaid, Ormar / Kuntz, Pascale (Eds.). *8<sup>èmes</sup> Journées Francophones Extraction et Gestion des Connaissances. Mesures de similarité sémantique*, Sophia 29. janvier - 1. Février 2008, pp. 29-38. Consultable sur :

<http://hal.univ-nantes.fr/docs/00/41/96/28/PDF/main.pdf> (article)

[www.egc.asso.fr/sdoc-94-egc08\\_atelier\\_mss.pdf](http://www.egc.asso.fr/sdoc-94-egc08_atelier_mss.pdf) (livre de conférence)

**Zargayouna**, Haifa/ Salotti, Sylvie (2004). “Mesure de similarité sémantique pour l'indexation de documents semi-structurés”. In : *12<sup>ème</sup> Atelier de Raisonnement à partir de Cas*, Ingénierie des Connaissances, Mars 2004, 10 pages. Consultable sur : <http://www-lipn.univ-paris13.fr/seminaires/AtelierRaPC/Articles/haifa.pdf>

## 7. Ressources numériques et manuels de thésauri

### a) Manuels de thésauri

**Carlisle**, Phil (2011). *Project Documentation of the European Heritage Network. Guidance Manual of the HEREIN Thesaurus Editing Software*, Ressource numérique dans la version du 29. Mars 2011, 27 pages, Consultable sur:

[http://www.coe.int/t/dg4/cultureheritage/heritage/herein/THES-GManual-2011\\_en.pdf](http://www.coe.int/t/dg4/cultureheritage/heritage/herein/THES-GManual-2011_en.pdf)

**Fillmore**, Charles J. / Ruppenhofer, Josef / Petruck, Miriam R.L. / Johnson, Christopher R./ Wood, Esther J. / Urban, Margaret / Baker Collin F. (2001). *The FrameNet Project: Tools for Lexicon Building. Version 0.7*, Ressource numérique dans la version du 13. Juillet 2001. Consultable sur :

<https://framenet.icsi.berkeley.edu/fndrupal/node/5347>

**Gouvernement du Québec** (2011). *Politique d'édition du Thesaurus de l'activité gouvernementale (TAG)*, Ressource numérique dans la version du 4. Juin 2011, 9 pages. Consultable sur :

[http://www.thesaurus.gouv.qc.ca/tag/acc\\_poltq\\_editn.do](http://www.thesaurus.gouv.qc.ca/tag/acc_poltq_editn.do)

**Gouvernement du Canada** (2009). *Thésaurus des sujets de base du gouvernement du Canada*, Ressource numérique dans la version du 3. Décembre 2009. Consultable sur :

[www.thesaurus.gc.ca](http://www.thesaurus.gc.ca) et sur [www.thesaurus.gouv.qc.ca/tag/accpoltq\\_editn.doc](http://www.thesaurus.gouv.qc.ca/tag/accpoltq_editn.doc)

**UQAM** (2009). Politiques d'indexation et de gestion du vocabulaire pour le Thésaurus du Service des bibliothèques de l'UQAM (RASUQAM), Ressource numérique dans la version du 8. Février 2006. Consultable sur :

<http://www.bibliotheques.uqam.ca/services-techniques/politiques-d%E2%80%99indexation-et-de-gestion-du-vocabulaire>

#### b) Ressources numériques

#### **Édition numérique du *Dictionnaire logique* (1882) :**

Blanc, Élie. *Dictionnaire Logique de la Langue Française ou Classification naturelle et philosophique des mots, des idées et des choses. Ouvrage pouvant servir de complément aux Dictionnaires et Encyclopédies alphabétiques*. Lyon - Vitte - Perrussel - Briday - Paris, 1882. Consultable sur :

<http://projects.chass.utoronto.ca/langueXIX/cabanes/index.html>

#### **Édition numérique du *Thesaurus of English Words* (1911) :**

Roget, Peter Mark. *Thesaurus of English Words and Phrases*, 4<sup>th</sup> edition, Londres 1911. Consultable sur le site du projet ARTFL - American and French research on the Treasury of the French Language :

<http://machaut.uchicago.edu/rogets>

#### **Dictionnaire numérique *Sémanticlopédie*, consultable sur :**

<http://www.semantique-gdr.net/>

L'article « Antonymie » de Pierre Amisli est consultable sur :

<http://www.semantique-gdr.net/dico/index.php/Antonymie>

L'article « Réseaux sémantiques » de Patrick Saint-Dizier est consultable sur :

[http://www.semantique-gdr.net/dico/index.php/Réseaux\\_sémantiques](http://www.semantique-gdr.net/dico/index.php/Réseaux_sémantiques)

**Site pour la création des lattices basés sur la consultation électronique de Roget's Thesaurus of English Words.**

Consultable sur : <http://www.ketlab.org.uk/roget.html>

**Site officiel de Uta Priss sur l'analyse formelle des concepts,**

Consultable sur : <http://www.upriss.org.uk/fca/fca.html>

**Site officiel du Centre nationales de Ressources textuelles et lexicales (CNRTL) avec diverses dictionnaires**

Consultable sur : <http://www.cnrtl.fr/>

### **Wikipédia**

Article « treillis », consultable sur : [www.wikipédia.fr/treillis](http://www.wikipédia.fr/treillis)

Article « guerre de cents ans », consultable sur : [www.wikipédia.fr/guerre-de-cent-ans](http://www.wikipédia.fr/guerre-de-cent-ans)

Article « château », consultable sur : [www.wikipédia.fr/chateau](http://www.wikipédia.fr/chateau)

## **Eidesstattliche Erklärung**

Hiermit versichere ich, **Sylvie Marie Magdelaine Reusch**,

dass ich die wissenschaftliche Prüfungsarbeit für die Erste Staatsprüfung für das Lehramt an Gymnasien selbstständig ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Diese Erklärung schließt auch die im Internet zugänglichen Daten ein. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind unter Angabe der Quellen der Entlehnung kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in gleicher oder anderer Form an irgendeiner Stelle als Prüfungsleistung vorgelegt worden.

\_\_\_\_\_

*Ort, Datum*

\_\_\_\_\_

*Unterschrift*